

The EUMETSAT  
Network of  
Satellite Application  
Facilities



**NWC SAF**

Support to Nowcasting and  
Very Short Range Forecasting

# Scientific and Validation Report for the Clear Air Product Processor of the NWC/GEO


NWC/CDOP2/GEO/AEMET/SCI/VR/ClearAir, Issue 1, Rev.0

15 October 2016

*Applicable to GEO-iSHAI v3.0 (NWC-031)*

## REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	Miguel A. Martinez (AEMET)		15 October 2016
Reviewed by	Xavier Calbet		15 October 2016
Authorised by	Pilar Ripodas SAFNWC Project Manager		15 October 2016

	<p>Scientific and Validation Report for the Clear Air Product Processor of the NWC/GEO</p>	<p><b>Code:</b> NWC/CDOP2/GEO/AEMET/SCI/VR/ClearAir  <b>Issue:</b> 1.0 <b>Date:</b> 15 October 2016  <b>File:</b> NWC-CDOP2-GEO-AEMET-SCI-VR-ClearAir_v1.0.docx  <b>Page:</b> 3/48</p>
---	--	--

## DOCUMENT CHANGE RECORD

Version	Date	Pages	CHANGE(S)
1.0d	15 November 2015	40	Draft version of the Scientific and Validation Report for the NWC/GEO Clear Air Product Processor GEO-iSHAIv3.0
1.0	15 October 2016	48	Updated with STRR changes and renamed GEO-TqPH as GEO-iSHAI

## TABLE OF CONTENTS

<b>1.</b>	<b>INTRODUCTION.....</b>	<b>8</b>
1.1	<i>SCOPE OF THE DOCUMENT .....</i>	8
1.2	<i>SOFTWARE VERSION IDENTIFICATION.....</i>	8
1.3	<i>GLOSSARY .....</i>	8
1.4	<i>REFERENCES .....</i>	11
1.4.1	<i>NWC SAF Applicable Documents .....</i>	11
1.4.2	<i>Reference Documents .....</i>	11
<b>2.</b>	<b>GEO-ISHAI VALIDATION DATASET .....</b>	<b>12</b>
2.1	<i>DESCRIPTION OF FILES USED .....</i>	12
2.2	<i>DESCRIPTION OF PROCESS TO GENERATE .....</i>	13
<b>3.</b>	<b>VALIDATION RESULTS.....</b>	<b>18</b>
3.1	<i>DISTANCE BETWEEN SEVIRI AND SYNTHETIC BTs AT DIFFERENT STEPS.....</i>	18
3.2	<i>ANALYSIS OF THE PERFORMANCE AT DIFFERENT VERTICAL LEVELS.....</i>	21
3.3	<i>2D DIMENSIONAL HISTOGRAMS OF GEO iSHAI PARAMETERS.....</i>	24
3.4	<i>SPATIAL ANALYSIS OF GEO iSHAI PARAMETERS .....</i>	29
3.5	<i>STATISTICAL SUMMARY OF LPW AND STABILITY INDICES.....</i>	33
3.6	<i>VALIDATION OF GEO-iSHAI TOZ: TOTAL OZONE.....</i>	37
3.7	<i>VALIDATION OF GEO-iSHAI SKT: SKIN TEMPERATURE.....</i>	40
<b>4.</b>	<b>CONCLUSIONS .....</b>	<b>43</b>
<b>5.</b>	<b>ANNEX I: VALIDATION STATISTICS OVER EUROPE REGION.....</b>	<b>45</b>

## List of Tables and Figures

Table 1: List of Applicable Documents. ....	11
Table 2: List of Referenced Documents.....	11
Table 3: <b>BT_RTTOV case:</b> Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ...	33
Table 4: <b>BT_SEVIRI case:</b> Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ...	34
Table 5: <b>BT_RTTOV case:</b> Statistical parameters for Lifted Index (LI), Showalter Index (SHW) and K Index (KI) parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ....	36
Table 6: <b>BT_SEVIRI case:</b> Statistical parameters for Lifted Index (LI), Showalter Index (SHW) and K Index (KI) parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ....	36
Table 7: <b>BT_RTTOV case:</b> Statistical parameters for Total Ozone (TOZ) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ...	39
Table 8: <b>BT_SEVIRI case:</b> Statistical parameters for Total Ozone (TOZ) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ...	40
Table 9: <b>BT_RTTOV case:</b> Statistical parameters for Skin Temperature (SKT) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ....	42
Table 10: <b>BT_SEVIRI case:</b> Statistical parameters for Skin Temperature (SKT) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels. ....	42
Table 11: Summary of the GEO iSHAI statistical parameters in 2013 for odd pixels of GEO iSHAI validation dataset using as input to GEO iSHAI actual SEVIRI BTs bias corrected.....	44
Table 12: Statistical accuracy values defined in the Product Requirement Table.....	44
Table 13: Statistical parameters for BL, ML, HL and TPW parameters over land Europe validation points in year 2013 for odd pixels dataset. Blue column) Green columns)BT_RTTOV case, light yellow columns) BT_SEVIRI case, red columns) uncorrected bias BT BT_SEVIRI case.....	48
Figure 1: Predefined set of 13001 validation points used in validation datasets. Grid network of 1° x 1° plus Radiosonde Stations (red crosses).....	13
Figure 2: Generation of the records for adding to GEO iSHAI validation dataset from one image. ...	15
Figure 3: <b>BT_SEVIRI case:</b> Spatial distribution of mean BT_distance (top) and BT_RMS(bottom) between real SEVIRI BTs and ECMWF analysis synthetic BTs at different step of GEO iSHAI when using as input bias corrected SEVIRI BTs. Left) forecast t+12 synthetic BTs, middle) synthetic BTs after FG step and right) using RTTOV BTs after physical retrieval step. Odd pixels on NWP-Hyb in year 2013. ....	19
Figure 4: <b>BT_RTTOV case:</b> Same that Fig. 3 but synthetic RTTOV BTs from ECMWF analysis are used as input to GEO iSHAI. Spatial distribution of mean BT_distance (top) and BT_RMS (bottom) between synthetic BTs from ECMWF (t+00) and synthetic BTs at different step of GEO iSHAI. Left) forecast t+12 synthetic BTs, middle) synthetic BTs after FG step and right) using RTTOV BTs after physical retrieval step. Odd pixels on NWP-Hyb in year 2013.....	19

- Figure 5: Histogram of top) BT\_distance (distance in IR10.8, IR12.0, WV6.2, WV7.3 and IR13.4 channels) and bottom) BT\_RMS (distance in absorption channels) at different steps of the GEO iSHAI over odd pixels in 2013. Left) BT\_RTTOV case, right) BT\_SEVIRI case. .... 20
- Figure 6: RMSE of q profiles (ppmv) at different steps compared with ECMWF analysis (t+00) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of q over sea pixels, bottom) RMSE of q over land pixels. Odd pixels in 2013..... 22
- Figure 7: RMSE of q profiles (ppmv) at different steps compared with ECMWF analysis (t+00) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of q over sea pixels, bottom) RMSE of q over land pixels. Odd pixels in 2013 on pixel with latitude greater than 36° N ..... 23
- Figure 8: **BT\_RTTOV** case: LPW and TPW 2D histograms over **sea** validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input BT\_RTTOV (t+00), right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles. .... 25
- Figure 9: **BT\_RTTOV** case: LPW and TPW 2D histograms over **land** validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input BT\_RTTOV (t+00), right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles. .... 26
- Figure 10: **BT\_SEVIRI** case: LPW and TPW 2D histograms over **sea** validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 27
- Figure 11: **BT\_SEVIRI** case: LPW and TPW 2D histograms over **land** validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles..... 28
- Figure 12: **BT\_RTTOV** case: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles. .... 29
- Figure 13: **BT\_RTTOV** case: Same that Figure 12 but relative RMSE instead of RMSE. .... 30
- Figure 14: **BT\_SEVIRI** case: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 31



- Figure 15: **BT\_SEVIRI** case: Same that Figure 14 but relative RMSE instead of RMSE. .... 32
- Figure 16: RMSE of ozone profiles (ppmv) at different steps compared with ECMWF analysis (t+00) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of ozone over sea pixels, bottom) RMSE of ozone over land pixels. Odd pixels in 2013..... 37
- Figure 17: Spatial distribution of the TOZ RMSE over validation points in 2013 for odd pixels dataset. (top) BT\_RTTOV case (bottom) BT\_SEVIRI case. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), right) GEO-TOZ RMSE calculated after non-linear regression step profile. In all case the ground truth are TOZ calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 38
- Figure 18: **BT\_RTTOV** case TOZ 2D histograms over odd pixels of year 2013. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background t+12ECMWF hybrid GRIB), (right) GEO-TOZ RMSE calculated after non-linear regression step profile. .... 39
- Figure 19: **BT\_RTTOV** case TOZ 2D histograms over odd pixels of year 2013. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background t+12ECMWF hybrid GRIB), (right) GEO-TOZ RMSE calculated after non-linear regression step profile. .... 39
- Figure 20: **BT\_RTTOV** case SKT 2D histograms over odd pixels of year 2013. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) SKT RMSE calculated after FG step profile, right) SKT RMSE calculated after physical retrieval step profile. In all case the ground truth is SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 40
- Figure 21: **BT\_SEVIRI** case SKT 2D histograms over odd pixels of year 2013. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) SKT RMSE calculated after FG step profile, right) SKT RMSE calculated after physical retrieval step profile. In all case the ground truth is SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 41
- Figure 22: Spatial distribution of the SKT RMSE over validation points in 2013 for odd pixels dataset. (top) BT\_RTTOV case (bottom) BT\_SEVIRI case. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), (centre) SKT RMSE calculated after FG non-linear regression step, (right) GEO-SKT RMSE calculated after physical retrieval step. In all case the ground truth are SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles. .... 41
- Figure 23: **BT\_RTTOV** case on Europe region: LPW and TPW 2D histograms over **land** on Europe region validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles. .... 46
- Figure 24: **BT\_SEVIRI** case on Europe region: LPW and TPW 2D histograms over **land** on Europe region validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles. .... 47
- Figure 25: Example of GEO-iSHAI diff\_ML and diff\_TPW from 12 UTC on 15 July 2015 produced from SEVIRI on MSG-3. Red pixels (large error in ML parameter) in the neighbourhood of cloudy pixels are caused by undetected clouds or cloud contamination..... 48

## 1. INTRODUCTION

The Eumetsat “Satellite Application Facilities” (SAF) are dedicated centres of excellence for processing satellite data, and form an integral part of the distributed EUMETSAT Application Ground Segment (<http://www.eumetsat.int>).

This documentation is provided by the SAF on Support to Nowcasting and Very Short Range Forecasting, hereafter NWC SAF. The main objective of NWC SAF is to provide, further develop and maintain software packages to be used for Nowcasting applications of operational meteorological satellite data by National Meteorological Services. More information can be found at the NWC SAF webpage, <http://www.nwcsaf.org>. This document is applicable to the NWC SAF processing package for geostationary meteorological satellites, NWC/GEO.

### 1.1 SCOPE OF THE DOCUMENT

The purpose of this document is to present the Scientific Validation Results for the Clear Air Product Processor of the NWC/GEO package release 2016. The name of Clear Air Product Processor is GEO-iSHAI (imagery Satellite Humidity And Instability).

All GEO-iSHAI coefficients and the validation figures have been calculated from the GEO iSHAI validation datasets using temporally and spatially collocated real SEVIRI and synthetic RTTOV brightness temperatures (calculated using profiles from ECMWF GRIB files on hybrid levels).

The scientific validation for GEO-iSHAI version 2016 outputs has been mainly based on the validation of the GEO iSHAI parameters using as input synthetic RTTOV SEVIRI brightness temperatures and real bias corrected SEVIRI brightness temperatures. Within this document, from now on GEO-iSHAI will only be referring to version 2016.

The case with synthetic BTs inputs is used to draw the main characteristics of the retrievals. The case using as input real bias corrected SEVIRI BTs is used to show the deviation from the theoretical one in operational conditions and to advice users on the limitation of the algorithm. To be able to make global validation and given that ECMWF analyses are of high quality, in this report these fields are taken as ground truth parameters.

### 1.2 SOFTWARE VERSION IDENTIFICATION

The validation results presented in this document apply to the GEO iSHAI version 3.0 product. This GEO-iSHAI version is included in the release 2016 of NWC/GEO software package.

### 1.3 GLOSSARY

Please refer to the “Nowcasting SAF Glossary” document in the NWC SAF web for a wider glossary and a complete list of acronyms for the NWC SAF project.

ABI	Advanced Baseline Imager
AEMET	Agencia Estatal de Meteorología Meteorology State Agency (Spain)
AHI	Advanced Himawari Imager
ASCII	American Standard Code for Information and Interchange
ATBD	Algorithm Theoretical Basis Document
BL	Precipitable water in low layer ( $P_{\text{sfc}} - 850 \text{ hPa}$ )
BT	Brightness Temperature



CDOP (CDOP-1)	Continuous Development and Operations Phase (1)
CDOP-2	Continuous Development and Operations Phase 2
CDOP-3	Continuous Development and Operations Phase 3
CF	NetCDF Climate and Forecast (CF) Metadata Conventions
CIMSS	Cooperative Institute for Meteorological Satellite Studies (USA)
CMa	Cloud Mask
COTS	Commercial-Off-The-Shelf
CPU	Central Processor Unit
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-range Weather Forecasts
EOF	Empirical Orthogonal Function
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCI	Flexible Combined Imager (MTG)
FG	First Guess
FOV	Field Of View
FOR	Field Of Regard
GEO	Geostationary Satellites
GEO-CMa	GEO Cloud Mask and Cloud Amount
GEO-iSHAI	GEO imagery Satellite Humidity And Instability
GRIB	Gridded Information in Binary Form
HDF5	Hierarchical Data format version 5
HL	Precipitable water in High Layer (500 – 0 hPa)
hPa	Hecto Pascal
HRIT	High Rate Image Transmission
IDL	Interactive Data Language
IR	InfraRed
IREMIS	InfraRed Emissivity
IRS	Infrared Sounder (MTG)
iSHAI	imagery Satellite Humidity And Instability
K	Kelvin
KI	K-Index
km	kilometre
LI	Lifted Index
LPW	Layer Precipitable Water
LST	Land Surface Temperature
MARS	ECMWF Meteorological Archive and Retrieval Facility
McIDAS	Man Computer Interactive Data Access System
ML	Precipitable water in Medium Layer (850 – 500 hPa)

MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
MTG-FCI	Meteosat Third Generation Flexible Combined Imager
MTG-IRS	Meteosat Third Generation Infra Red Sounder
netCDF	Network Common Data Form
NRT	Near Real Time
NWC	Nowcasting
NWC/GEO	Geostationary part of the Nowcasting SAF
NWCLIB	Nowcasting Library
NWC SAF	Nowcasting SAF
NWP	Numerical Weather Prediction
NWP SAF	SAF for Numerical Weather Prediction
LPW	Layer Precipitable Water
PGE	Product Generation Element PGE01 Cloud Mask (GEO-CMa) Product Generator PGE13 SEVIRI Physical Retrieval (SPhR) Product Generator
PW	Precipitable Water
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVs
SAF	Satellite Application Facility
SEVIRI	Spinning Enhanced Visible InfraRed Imager
SG	Steering Group
SHAI	Satellite Humidity And Instability
SHW	Showalter Index
SKT	Skin Temperature
SST	Sea Surface Temperature
SW	Software
TOZ	Total ozone
TPW	Total Precipitable Water
TM	Task Manager
UM	User Manual
VR	Validation Report
VSA	Visiting Scientist Activities
WV	Water Vapour Channel

## 1.4 REFERENCES

### 1.4.1 NWC SAF Applicable Documents

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD.X]

For versioned references, subsequent amendments to, or revisions of, any of these publications do not apply. For unversioned references, the current edition of the document referred applies.

Current documentation can be found at the NWC SAF Helpdesk web: <http://www.nwcsaf.org>.

Ref.	Title	Code	Vers
[AD.1]	Proposal for the Second Continuous Development and Operations Phase (CDOP2)	NWC/CDOP2/MGT/AEMET/PRO	1.0
[AD.2]	NWC SAF CDOP-2 Project Plan	NWC/CDOP2/SAGF/AEMET/MGT/PP	1.9
[AD.3]	Configuration Management Plan for NWC SAF	NWC/CDOP2/SAF/AEMET/MGT/CMP	1.4
[AD.4]	NWC SAF Product Requirements Document	NWC/CDOP2/SAF/AEMET/MGT/PRD	1.9
[AD.5]	Interface Control Document for Internal and External Interfaces of the NWC/GEO	NWC/CDOP2/AEMET/SW/ICD/1	1.1
[AD.6]	Data Output Format for the NWC/GEO	NWC/CDOP2/AEMET/SW/DOF	1.1
[AD.7]	System Version Document for the NWC/GEO	NWC/CDOP2/AEMET/SW/SCVD	1.1

*Table 1: List of Applicable Documents.*

The reference documents contain useful information related to the subject of the project. These reference documents complement the applicable ones, and can be looked up to enhance the information included in this document if it is desired. They are referenced in this document in the form [RD.X]

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies.

### 1.4.2 Reference Documents

Ref.	Title	Code	Vers	Date
[RD.1]	Validation Report for "PGE13 SEVIRI Physical Retrieval" (SPhR- PGE13 v1.2)	SAF/NWC/CDOP2/INM/SCI/VR/11	1.0	15/02/12
[RD.2]	Scientific Report: improvements in "PGE13 SEVIRI Physical Retrieval Product (SPhR)" using as input ECMWF GRIB files on hybrid levels	SAF/NWC/CDOP2/INM/SCI/RP/02	1.0	15/07/13
[RD.3]	Algorithm Theoretical Basis Document for the Clear Air Product processor of the NWC/GEO	NWC/CDOP2/GEO/AEMET/SCI/ATBD/ClearAir	1.1	15/10/16
[RD.4]	User Manual for the Clear Air Product Processors of the NWC/GEO: Science Part	NWC/CDOP2/GEO/AEMET/SCI/UM/ClearAir	1.0	15/10/16

*Table 2: List of Referenced Documents*

## 2. GEO-ISHAI VALIDATION DATASET

### 2.1 DESCRIPTION OF FILES USED

In order to build the GEO-iSHAI datasets for training and validation purposes, data from MSG SEVIRI images and ECMWF GRIB files have been used. At the time of writing this report the input data files available for training, tuning and validation activities are:

#### From ECMWF fields:

- 00 Z and 12 Z runs
- analysis ( $t+00$  hours) and forecasts ( $t+12$  hours)
- region: NW corner at (65° N, 65° W) and SE corner at (65° S, 65° E)
- time period: from 31 December 2007 12 Z to present of each day during this period. The database is updated every month.
- horizontal resolution: 0.5° by 0.5°
- vertical resolution: two different vertical resolutions are used
  - o Hybrid levels (hereafter denoted as NWP-Hyb): The number is 91 or 137 levels depending on the ECMWF cycle.
  - o Fixed pressure levels (hereafter denoted as NWP-P): These GRIB files are needed only for PGE01 CMA processing. The pressure levels available on MARS are typically 15 or 25 synoptic levels (as example: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100 .. hPa)
- parameters: temperature (T), humidity (specific humidity [q] for NWP-Hyb and relative humidity [RH] for NWP-P files).

*Note: the NWP-P GRIB files are the ones used as input to the MSG NWCSAF software package.*

*Note: when the collocated records are written, the NWP( $t+12$ ) from previous 12 hour ECMWF run are collocated in the same record with the NWP( $t+00$ ). As example, the 01 January at 00UTC NPW( $t+00$ ) analysis profile is collocated in the same record than the one with the NWP( $t+12$ ) from 31 December 12 UTC.*

#### From MSG SEVIRI Observations:

- 00 Z and 12 Z slots
- region: frame of 3400 x 3400 IR pixels centred at subsatellite position (only pixels with satellite zenith angle lower than 70° )
- time period: from 1 January 2008 00 Z to present of each day during this period (continuous update)
- horizontal resolution: SEVIRI full resolution and MSG projection
- SEVIRI channels: All SEVIRI channels but HRVIS
- Currently in the validation datasets there are MSG data from MSG-1, MSG-2, MSG-3 and MSG-4. In this report the period selected is the year 2013 and for this reason MSG-3 is the satellite used when actual SEVIRI brightness temperatures (hereafter BTs) are used.

These are the dynamic information datasets used for the tuning and validation activities. Specific datasets used for different objectives are in part generated from them and descriptions are provided in the respective sections.

## 2.2 DESCRIPTION OF PROCESS TO GENERATE

To build one good validation and training dataset is a high priority and standing task for us. The process to build the GEO-iSHAI v2016 validation dataset is a heritage of the process used to build the PGE13 SPhR validation dataset; see [RD.1] or [RD.2].

The main idea is to generate the whole validation dataset using PGE01 CMA and a modified version of PGE13 SPhR from NWSAF/MSG v2013 software called PGE00Hyb which uses RTTOV-11.2 instead of RTTOV-9.3.

PGE13 SPhR, PGE00Hyb and GEO-iSHAI allow activating an option in their configuration files to write in optional binary files structures at clear pixels with real SEVIRI BTs together with the T, q and O<sub>3</sub> profiles, surface and ancillary parameters collocated spatially, temporally and vertically interpolated to the position and time of the clear SEVIRI pixels.

The use of RTTOV-11.2 implies that the GEO-iSHAI validation dataset is based now on profiles with 54 pressure levels; instead of the 43 pressure levels from RTTOV-9.3. In the near future the process will be migrated to NWC/GEO v2016 GEO-CMA and GEO-iSHAI. Thus, all the profiles used in the validation have similar characteristics to the profiles used and retrieved within the GEO-iSHAI v2016 execution in operational mode.

The 2013 year has been chosen as the reference period for the GEO-iSHAI validation dataset. The validation results obtained using as ground truth the ECMWF NWPHyb analysis ( $t+00$ ) profiles are presented here. Also the validation results using as input to GEO-iSHAI real SEVIRI BTs bias BT corrected are included in order to show the deviation from the theoretical ones and to advice users on the limitation of the algorithm.

To avoid using for validation the same records as the ones used for calculation of the GEOT-iSHAI 2016 version coefficients, the records with odd position in the complete dataset have been used for validation.

**Positions of GEO iSHAI validation dataset:** For GEO iSHAI parameters validation, a set of predefined positions of a 1° x 1° grid plus the RAOB stations positions have been chosen. The set contains 13001 points in the actual mask. The positions where validation is made can be seen in Figure 1.

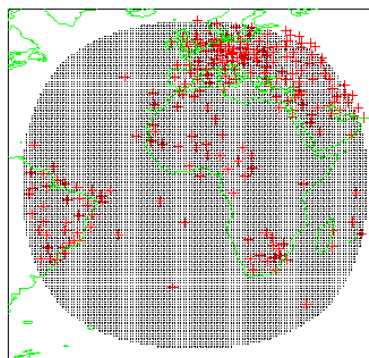


Figure 1: Predefined set of 13001 validation points used in validation datasets. Grid network of 1° x 1° plus Radiosonde Stations (red crosses).

**Process to build the GEO iSHAI validation dataset:** The actual process to build the validation and training dataset is the following:

a) Calculate Cloud Mask (PGE01 CMA): the cloud mask generation is the first step. The PGE01 program is first executed. The results of PGE01 v2013 program are HDF-5 files with the cloud mask located at \$SAFNWC/export/PGE01.

Please, note that for PGE01CMA the NWP GRIB files used as input need to be on fixed pressure levels whilst for this Validation Report all calculations come from hybrid ECMWF GRIB files (91 hybrid levels till June 2013 and 137 hybrid levels in rest of 2013). In this report all background NWP profiles have been downloaded from ECMWF.

*Note: In near future this step will be migrated to GEO-CMA release 2016 and netCDF files.*

b) Selection of clear validation locations by screening the cloud mask: the NWC SAF PGE01 CMA cloud mask file is overwritten with the multiplication of PGE01CMA mask with the 13001 validation position mask (1 for selected pixel and 0 in the rest). The use of this screened cloud mask speeds up the process because instead of executing the physical retrieval over several millions of clear air pixels it is executed only over the clear air pixels among the 13001 predefined positions.

The process to get the screened cloud mask is:

- the cloud mask matrix is read from the PGE01 CMA file
- the cloud mask matrix is multiplied with the 13001 validation positions mask (matrix with same dimensions of CMA with values: 1 for validation points and 0 for the rest of pixels) and the result is the screened cloud mask.
- This screened cloud mask is used to overwrite the PGE01\_CMA cloud mask HDF-5 file on `$SAFNWC/export/PGE01`.

c) To obtain the profiles from analysis ( $t+00$ ) and forecast ( $t+12$ ): the PGE00Hyb program is executed twice for each slot. The PGE00Hyb program calculates the profiles by interpolating the ECMWF fields from hybrid levels to 54 levels in the vertical and also in time and space. It also calls RTTOV-11.2 to calculate the synthetic BTs.

In the first PGE00Hyb execution: the screened cloud mask, the SEVIRI image and as background NWP-Hyb the  $t+12$  forecast ECMWF GRIB file from previous 12 hour ECMWF run (hereafter denoted as NWPH12) are used as inputs.

In the second PGE00Hyb execution: the screened cloud mask, the SEVIRI image and as background NWP the ECMW  $t+00$  analysis GRIB file (hereafter denoted as NWPH00) are used as inputs.

With the first execution it is read the (T, q, O<sub>3</sub>) profiles and some surface parameters (P<sub>sfc</sub>, T<sub>skin</sub>, etc), from the `$SAFNWC/tmp` binary files, at the clear air predefined positions from  $t+12$  hours ECMWF forecast.

In the second execution it is read the (T, q, O<sub>3</sub>) profiles and some surface parameters at the clear air predefined positions from ECMWF  $t+00$  analysis.

Together with the  $t+00$  and  $t+12$  profiles, ancillary data (as emissivities, longitude, latitude, zenith angle, etc) are also read from the `$SAFNWC/tmp` binary files. The process can be seen in the Figure 2. The result is one binary file by slot that can be easily read on IDL with the `restore` command. The mean number of retained clear pixels by slot is 4829.



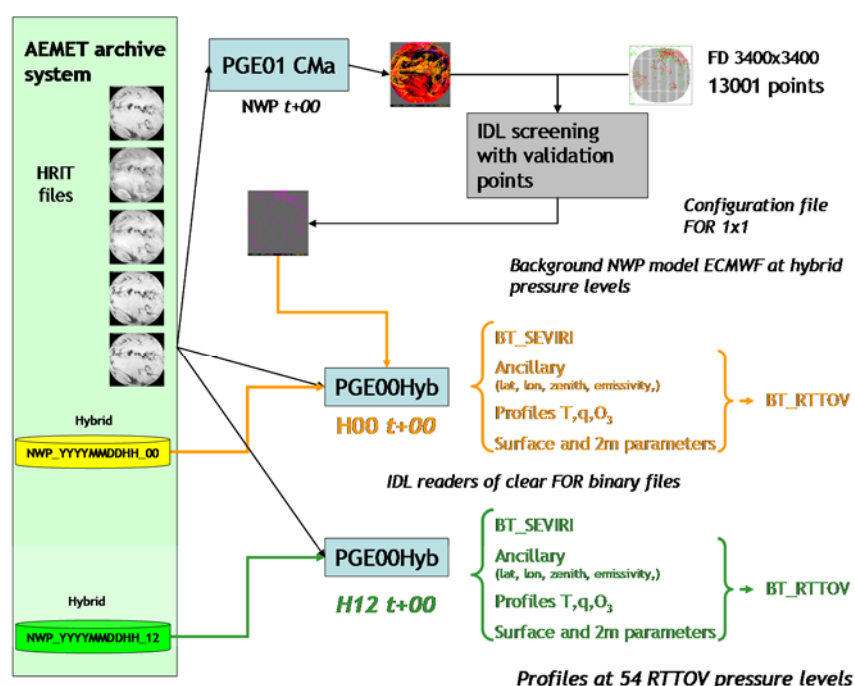


Figure 2: Generation of the records for adding to GEO iSHAI validation dataset from one image.

d) Joining the files for every slot on monthly files: In order to allow an easy management of the datasets, the slot binary files are joined in one file for every month. It is made with one IDL procedure. This monthly binary files are the base for the validation process since files on a monthly basis can be joined easily to build a wider period dataset.

e) Joining the monthly files on a period file: Once a period is selected for validation or training, one period binary file is generated joining the monthly files for the months in the period. It is made with one IDL procedure. In this validation report it has been used 1 out 2 (pixels with odd positions) clear pixels for the 2013 year. The number of pixels used here is 1,709,518 pixels. Other reason to use 1 out 2 of the 2013 clear pixels is that is needed 3.5 GB to store it in memory (the compressed file has a size of 2.5 GB).

f) Write binary file which can be used as input to the validation version of GEO-iSHAI: One array with selected parameters is written on a binary file in a format that will be used later with the validation version of GEO-iSHAI. This validation version processes the data record by record instead of processing a MSG image, as is done in the operational version.

g) Execute the validation version of the GEO-iSHAI software: to get a assessment of the new coefficients for the First-Guess (hereafter FG) regression and the physical retrieval steps of 2016 GEO-iSHAI, one validation version of the sources of PGE13Hyb to process PGE13 and iSHAI algorithm with the Hybrid coefficients and RTTOV-11.2 on record by record basis has been developed. Thus, it is possible to test the new FG regressions coefficients, Empirical Orthogonal Function (EOF) coefficients files,  $E^{-1}$  and  $B^{-1}$  matrices, avoiding the huge task to reprocess from the complete HRIT SEVIRI files and ECMWF GRIB files.

This also allows to choose use as input to the algorithm real SEVIRI BTs (bias corrected or uncorrected) or synthetic RTTOV SEVIRI BTs. The outputs of this validation program are the profiles after FG regression and/or physical retrieval steps using as inputs the hybrid profiles. This process allows testing new version of iSHAI software or new coefficients just over GEO-iSHAI validation dataset before implementing it in the GEO-iSHAI operational software. This process is a natural consequence from previous experiences.

h) Build the GEO iSHAI validation dataset: The outputs of the previous step are blended with the structures of the validation dataset (ancillary fields as emissivity values, land/sea mask, height, etc.) and one IDL binary file for restoring is generated.

**Structure of the records in the GEO iSHAI validation dataset**: After the execution of the previous steps the validation dataset for a period is one array (that could contains several millions of records) written on an IDL binary file. The array can be restored easily with the use of IDL command *restore*. Every record is one IDL structure with the following parameters or fields:

- **Ancillary**: longitude, latitude, emissivity values, etc.
- **Date**: day, year, hour, etc.
- **NWP from ECMWF analysis (t+00)**: ECMWF temperature and humidity profiles interpolated to the 54 RTTOV pressure levels interpolated vertically from the 91 or 137 hybrid levels,  $T_{skin}$ , pressure at surface, etc. from the analyses (t+00) ECMWF GRIB files. It will be used as the validation truth.
- **NWP from ECMWF forecast (t+12)**: ECMWF temperature and humidity profiles interpolated to the 54 RTTOV pressure levels interpolated vertically from the 91 or 137 hybrid levels,  $T_{skin}$ , pressure at surface, etc. from the previous run to the image ECMWF t+12 forecast (as example for image 20090101 at 00Z the t+12 forecast from 20081231 at 12 UTC ECMWF run is used).
- **BT\_SEVIRI(8)**: uncorrected BT from HRIT file.  $BT\_SEVIRI[IR3.9, WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]$ .
- **BT\_RTTOV(8) from NWPHyb(t+00)**: Synthetic BTs calculated using the RTTOV-11.2 with the analysis (t+00).  $H00.BT\_RTTOV[IR3.9, WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]$ .
- **BT\_RTTOV(8) from NWPHyb(t+12)**: Synthetic BTs calculated using the RTTOV-11.2 with the forecast (t+12).  $H12.BT\_RTTOV[IR3.9, WV6.2, WV7.3, IR10.8, IR8.7, IR9.7, IR12.0, IR13.4]$ .

These basic validation and training datasets have been used for several tasks. One of them consisting of splitting the records in the global dataset into records to generate the validation datasets (odd positions) and the training ones (even positions). All the 2016 version of GEO-iSHAI coefficients have been calculated with this dataset. For the GEO-iSHAI the changes with respect to PGE13 SPhR v2013 in the coefficients files are the following:

- **New FG non-linear regression coefficients**: 200 thousand profiles (half sea pixels and half land pixels) has been extracted from year 2013 training dataset with a random process in which the probability to be extracted to this reduced training dataset increases with the inverse to the frequency in the histogram of TPW. This reduced dataset with more uniform representativeness of profiles with different precipitable water have been used to train the hybrid FG regressions. The FG regressions coefficient file contains 76 regression coefficients for different local zenith angles ranging from 0 to 75 degrees and for every atmospheric parameter (54 for temperature profile, 54 for humidity profile and 1 for skin temperature). See GEO-iSHAI ATBD [RD.3] for details on the fields used in the FG regressions.
- **New temperature and log(q) EOF files**: profiles from analysis have been used in calculation of new EOFs for temperature and log(q) profiles. Thus, these new EOFs files are based on year 2013 ECMWF hybrid profiles of the training dataset.
- **New inverse of covariance matrix of the background error  $B^{-1}$** : the covariance matrix of the background error between t+12 and t+00 ECMWF profiles has been calculated using NWP-Hyb dataset from the 2013 year records. Thus, the  $B^{-1}$  matrix is based on ECMWF hybrid profiles from the whole MSG disk.
- **New inverse of matrix measurement error  $E^{-1}$** : the sea pixels of the 2013 year PGE13 training dataset are used to calculate the  $(BT_{SEVIRI} - BT_{RTTOV})$  matrix for channels [WV6.2, WV7.3, IR10.8, IR12.0, IR13.4]. Then, the diagonal of this matrix is retained and its inverse is calculated. The  $E^{-1}$  matrix is important since it controls the rate of conversion from  $(BT_{SEVIRI}$

–BT<sub>RTTOV</sub>) values to modifications on the (T, log(q)) profiles. For convenience in GEO-iSHAI algorithm (see section 2.2.6 of [RD-3]), the values in the E<sup>-1</sup> files are the square root.

To avoid some issues (as SEVIRI BT biases, emissivity, contamination by clouds, need of screening in the selection of the records, etc.) and to make the document more readable, the assessment of the performances for these new GEO-iSHAI coefficients files is shown first when GEO-iSHAI uses as input synthetic RTTOV brightness temperatures from ECMWF hybrid analysis (*t+00*) profiles; this experiment is denoted hereafter as **BT\_RTTOV case**. Then, the performances are compared with the ones using as input to GEO-iSHAI real bias corrected SEVIRI BTs; this experiment is denoted hereafter as **BT\_SEVIRI case**.

### 3. VALIDATION RESULTS

Results presented are organised as follows. In Section 3.1, the analysis of the error between the SEVIRI BTs and the synthetic RTTOV BTs at the different steps of the algorithm is first discussed. In section 3.2 an analysis of the vertical performance of the algorithm is made. In section 3.3 the analysis of 2D histograms is made. In Section 3.4 the analysis of spatial validation results are presented. In Section 3.5 the statistical values for LPW, TPW and stability indices are summarized in tables. In Section 3.6 and in Section 3.7 the statistical values for TOZ and SKT are shown respectively.

#### 3.1 DISTANCE BETWEEN SEVIRI AND SYNTHETIC BTs AT DIFFERENT STEPS

In order to check the added value of the successive steps of the GEO iSHAI algorithm, an inspection of the difference between the synthetic RTTOV BTs calculated using the profiles from the ECMWF analysis ( $t+00$ ) and used here as true profiles, versus the real bias corrected SEVIRI BTs and the synthetic RTTOV BTs at the different steps of the GEO iSHAI algorithm is made first. This inspection has been divided in two parts. First, a spatial analysis of the BT differences is made and later an analysis of the histogram of the BT error is done.

In this subsection, two parameters are checked. The first parameter is the distance between synthetic BTs obtained from profiles from the GEO iSHAI physical retrieval step and real BTs in all SEVIRI channels (IR10.8, IR12.0, WV6.2, WV7.3 and IR13.4). This is denoted as BT\_distance. The second one, BT\_RMS, is the same statistics as before, but calculated using non-window channels, i.e. the distance between BTs calculated from WV6.2, WV7.3 and IR13.4 channels.

Before the generation of the figures and the statistical results on this report, it has been implemented one outliers screening removing the 5% of pixels with the largest BT\_RMS and BT\_distance. The Figures before the screening are not shown in this report.

In Figure 3, real bias corrected SEVIRI BTs are used as input to the GEO iSHAI and it is denoted hereafter as BT\_SEVIRI case. The spatial distribution of the mean BT\_distance and mean BT\_RMS are shown at the three main steps on the GEO iSHAI algorithm. In BT\_SEVIRI case, **H12** step means that to calculate BT\_distance and BT\_RMS real bias corrected SEVIRI BTs and synthetic forecast  $t+12$  BTs have been used. **FG** step means that to calculate BT\_distance and BT\_RMS real bias corrected SEVIRI BTs and profiles after FG regression have been used. **I1** step means that to calculate BT\_distance and BT\_RMS real bias corrected SEVIRI BTs and profiles after physical retrieval routine have been used.

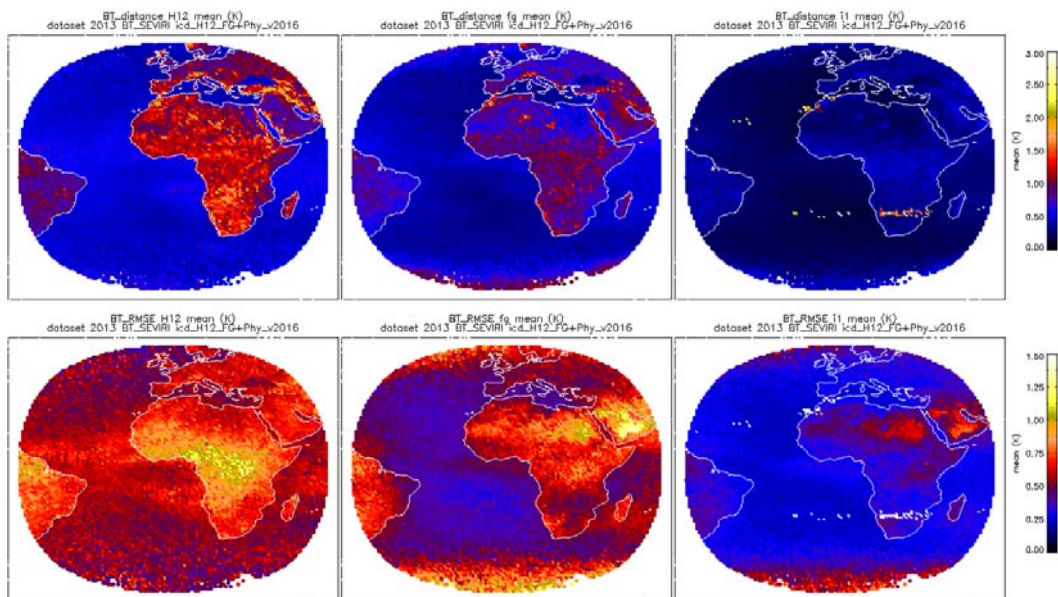
It can be seen in Figure 3 that the FG regression and the physical retrieval steps significantly reduces both BT\_distance and BT\_RMS. That means that non-linear regression and physical retrieval effectively modify the profiles in correct direction to get a convergence of synthetic and real BTs. The regions with largest errors and residuals are located over land areas.

Figure 4 is similar to Figure 3, but in Figure 4 synthetic RTTOV BTs from the ECMWF analysis ( $t+00$ ) (hereafter BT\_RTTOV case) are used as input to the GEO iSHAI module (instead of real bias corrected SEVIRI BTs). It can be seen that the GEO iSHAI in Figure 4 has similar behaviour than in Figure 3 but without the regions with larger errors that appear in BT\_SEVIRI case (when real bias corrected SEVIRI BTs are the input).

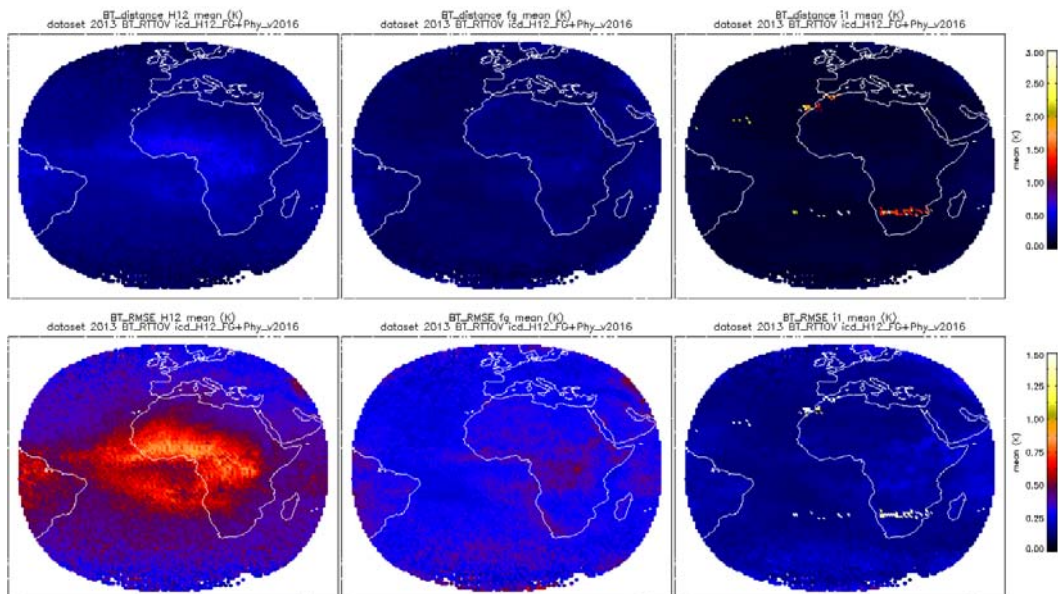
The larger residuals on Figure 3 compared to Figure 4 are associated to several issues, such as uncertainties on emissivity fields, contamination with clouds on some pixels (not well filtered by PGE01 Cloud Mask), errors in the radiative transfer model RTTOV, noise of real SEVIRI image, etc. On Figure 3, the larger values of BT\_distance are also likely due to errors between SEVIRI BT at window channels and the ones calculated from ECMWF analysis due to the fact that skin



temperature on ECMWF analysis does not represent real skin temperature over desert regions, mountains, etc.



**Figure 3: *BT\_SEVIRI* case:** Spatial distribution of mean *BT\_distance* (top) and *BT\_RMS*(bottom) between real *SEVIRI* BTs and ECMWF analysis synthetic BTs at different step of GEO iSHAI when using as input bias corrected *SEVIRI* BTs. Left) forecast  $t+12$  synthetic BTs, middle) synthetic BTs after FG step and right) using *RTTOV* BTs after physical retrieval step. Odd pixels on NWP-Hyb in year 2013.



**Figure 4: *BT\_RTTOV* case:** Same that Fig. 3 but synthetic *RTTOV* BTs from ECMWF analysis are used as input to GEO iSHAI. Spatial distribution of mean *BT\_distance* (top) and *BT\_RMS* (bottom) between synthetic BTs from ECMWF ( $t+00$ ) and synthetic BTs at different step of GEO iSHAI. Left) forecast  $t+12$  synthetic BTs, middle) synthetic BTs after FG step and right) using *RTTOV* BTs after physical retrieval step. Odd pixels on NWP-Hyb in year 2013.

It can be seen in Figure 3 and 4 that the FG regression and the physical retrieval steps significantly reduces both *BT\_distance* and *BT\_RMS*. As conclusion of the analysis of Figures 3 and 4, the physical retrieval algorithm implemented in the GEO iSHAI algorithm works fine and the retrieved

(T, q) profiles reduces significantly the distance between the bias corrected SEVIRI BTs and the synthetic RTTOV BTs.

On the other hand, the analysis of GEO iSHAI algorithm steps with synthetic RTTOV as input can be representative of the performance of the GEO iSHAI algorithm avoiding the issues related to real measurements mentioned in the previous paragraph. For these reasons, and to avoid the need of additional screening filters, especially hard to make over land, the validation on the next sections is made always showing first the statistical values from the synthetic BT\_RTTOV case.

Not shown here, similar figures using real uncorrected bias SEVIRI BTs have been calculated; the performance in the case of actual uncorrected bias BT is worse than the performance in the case of corrected bias BT; this fact is another justification on the need of a good and updated BT bias correction.

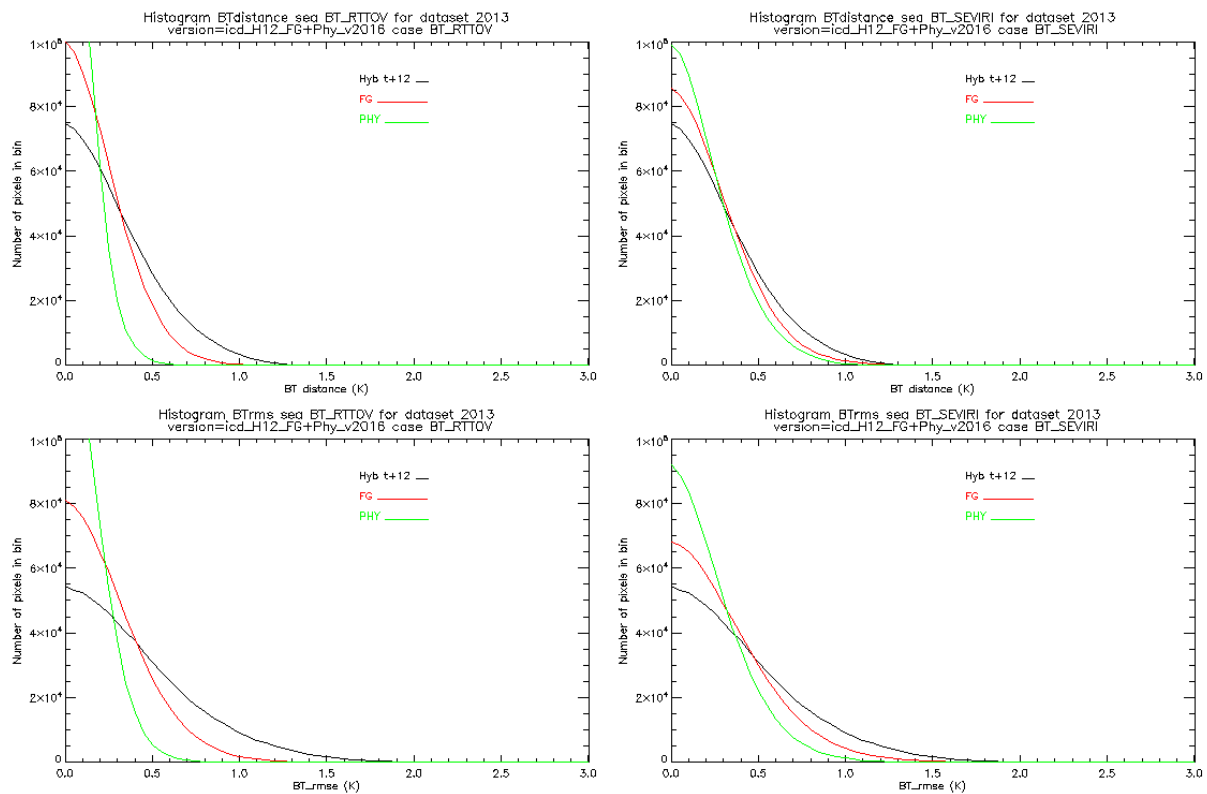


Figure 5: Histogram of top) BT\_distance (distance in IR10.8, IR12.0, WV6.2, WV7.3 and IR13.4channels) and bottom) BT\_RMS (distance in absorption channels) at different steps of the GEO iSHAI over odd pixels in 2013. Left) BT\_RTTOV case, right) BT\_SEVIRI case.

The histograms with BT\_distance and with BT\_RMS for the BT\_RTTOV case and the BT\_SEVIRI case at the different steps of the GEO iSHAI module are shown in Figure 5. It can be seen that the physical retrieval step reduces significantly the number of profiles with BT\_distance and BT\_RMS greater than 0.6. Also, the FG step reduces both BT\_distance and BT\_RMS. The BT\_RMS has been verified because in the GEO iSHAI code if the BT\_RMS in the pixel is greater than a threshold, then the physical retrieval module is applied. The analysis of the histograms of the error and BT\_RMS can be used to select the optimal values for the configurable parameters BT\_RMS\_THRESHOLD and MAX\_RESIDUAL (these parameters are read from the ASCII GEO-iSHAI configuration file and an explanation of the impact in the selection in these parameters can be found in the User Manual [RD-4]).



### 3.2 ANALYSIS OF THE PERFORMANCE AT DIFFERENT VERTICAL LEVELS

In Figure 6, the RMSEs between the  $q$  profiles after several steps in the GEO iSHAI for the full disc dataset at the 54 RTTOV levels have been represented.

The profiles of the ECMWF analysis ( $t+00$ ) from NWP-Hyb datasets (profiles from 91 or 137 hybrid levels interpolated to the 54 RTTOV level) have been considered as the truth. Only odd pixels have been used here in the vertical profile analysis in order not to use the same ones as the ones used to calculate the version 2016 GEO iSHAI coefficients.

The statistical values for the specific humidity at mid levels show better performance for the FG regression and the physical retrieval steps than the background NWP model (ECMWF GRIB files on hybrid levels from  $t+12$  forecast). Likely due to the added value of the WV SEVIRI channels, the reduction in the RMSE at these levels indicates that the GEO iSHAI slightly improves the  $q$  profile from background NWP.

The performance is better over sea pixels. The worse performance over land can be due to a combination of all sources of errors that can affect the calculation of the synthetic BT (uncertainties in the emissivity atlas, errors due to contamination with clouds on some pixels (not well filtered by the Cloud Mask), errors in the radiative transfer model RTTOV, noise of actual SEVIRI image, etc.)

Several experiments, not shown here, to compare the GEO iSHAI performance have been made. The first experiment is to suppress the FG regression step and use directly as First Guess the  $t+12$  forecast profiles from ECMWF GRIB files on hybrid levels; but the performance of GEO iSHAI with FG regression step is slightly better.

These experiments and the faster executions of GEO iSHAI with FG regression step (avoiding the execution of the physical retrieval step on all clear FOR) have inclined us to maintain the FG regression step on the GEO iSHAI algorithm.

It can be seen in 2012 Validation Reports [RD.1] that the use of ECMWF with 15 fixed pressure levels profiles as input to the PGE13 SPhR created several “peaks” and irregularities in the RMSE and bias vertical distribution centred at the 15 fixed pressure levels. The comparison of Figure 6 of this report with Figure 11 of the 2012 Validation Report [RD.1] is another reason to strongly recommend the use of GRIB files with the maximum number of vertical levels as possible as the background NWP input to GEO iSHAI. For this reason the performance of GEO iSHAI processor with GRIB files on hybrid levels as background NWP input is the best possible.

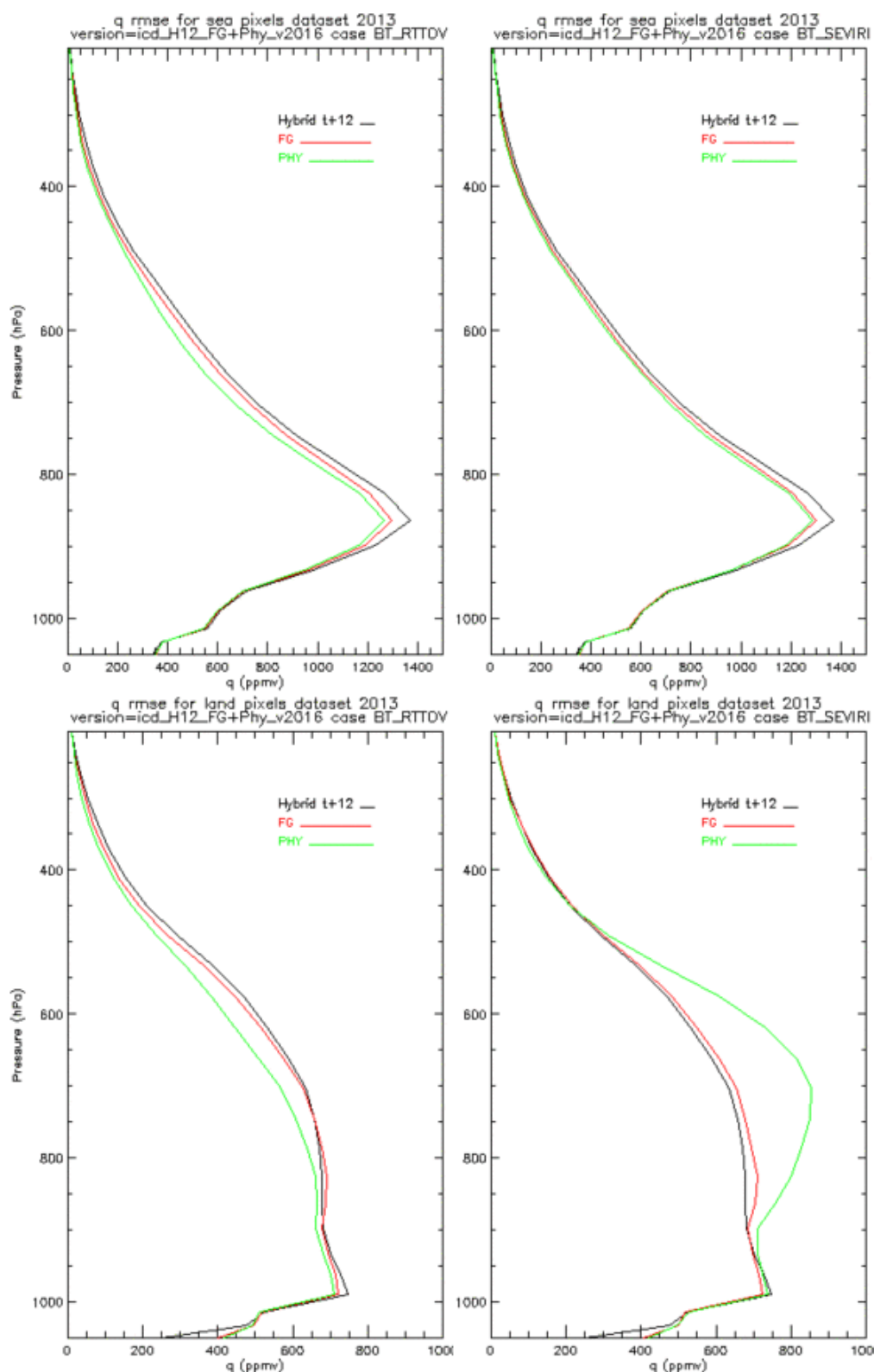


Figure 6: RMSE of  $q$  profiles (ppmv) at different steps compared with ECMWF analysis ( $t+00$ ) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of  $q$  over sea pixels, bottom) RMSE of  $q$  over land pixels. Odd pixels in 2013

The NWC/GEO package version 2016 will continue with the use of the fixed pressure levels as background NWP input to GEO iSHAI, since the current version and the 2016 version of NWC SAF library only allows reading GRIB files on fixed pressure. GEO-iSHAI is just one advance of

future versions of NWC/GEO processing. The details to activate GEO-iSHAI on hybrid levels instead to the regular GEO iSHAI on P levels can be found in the GEO-iSHAI User Manual [RD.4].

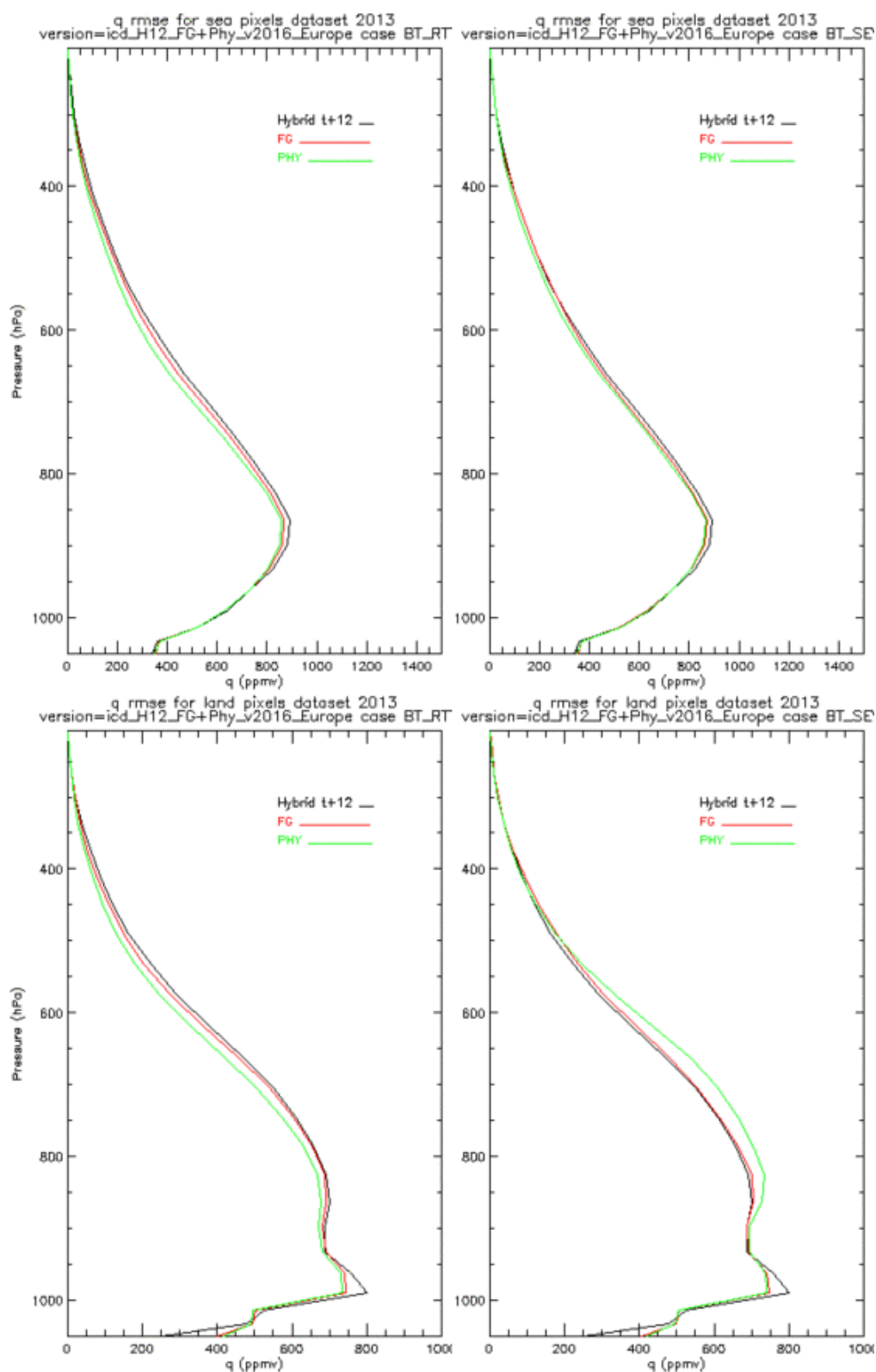


Figure 7: RMSE of  $q$  profiles (ppmv) at different steps compared with ECMWF analysis ( $t+00$ ) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of  $q$  over sea pixels, bottom) RMSE of  $q$  over land pixels. Odd pixels in 2013 on pixel with latitude greater than  $36^\circ$  N

In order to show that the source of the increase of  $q$  rmse on land SEVIRI case is associated to the high  $q$  rmse error in Equatorial region (see Figure 14 in section 3.4), in the Figure 7 the profiles of  $q$  rmse but for European latitude (considered as latitude greater than  $36^{\circ}\text{N}$ ) are shown. It can be seen the strong reduction in  $q$  rmse in the land SEVIRI case in the layer 900-600 hPa.

### 3.3 2D DIMENSIONAL HISTOGRAMS OF GEO iSHAI PARAMETERS

To avoid multiplying the number of Figures, only the two dimensional histograms for each one of the LPW and TPW parameters calculated from the retrieved profiles at different steps are presented here. It has been used always as truth the ECMWF analysis denoted here as NWP-Hyb ( $t+00$ ) profiles. In Figure 8 for sea pixels and in Figure 9 for land pixels for BT\_RTTOV case. Same Figures but for BT\_SEVIRI case can be seen in Figures 10 and 11. The statistical values (RMSE, bias and correlation) that appear in the 2D histograms are also written at the end of this report in the tables of the Section 3.5 for a more comfortable read and comparison.

Take note that GEO-iSHAI BL is the precipitable water in a layer between  $P_{\text{sfc}}$  to 850 hPa. GEO-iSHAI ML is the precipitable water in a layer between 850 hPa to 500 hPa. GEO iSHAI HL is the precipitable water in a layer between 500 hPa to 0.1 hPa. GEO iSHAI TPW is the total precipitable water i.e the precipitable water in a layer between  $P_{\text{sfc}}$  to 0.1 hPa.

It can be seen in Figures 8 to 11 that statistical values of the GEO-iSHAI parameters reproduce the performance suggested by the vertical analysis from Figure 6 and 7. The parameters with the largest added value are ML and HL parameters; this fact is due to the WV channels.

Other important result is that the 2D histograms of the GEO-iSHAI parameters show no significant bias and it is not needed any post processing correction.

In the case of BT\_SEVIRI case over land some spread in the 2D histograms shows up; the cause of this spread could be due to the factors mentioned before (cloud contaminated pixels, uncertainties in the emissivity atlases, disagreements in the skin temperature from the ECMWF, etc).

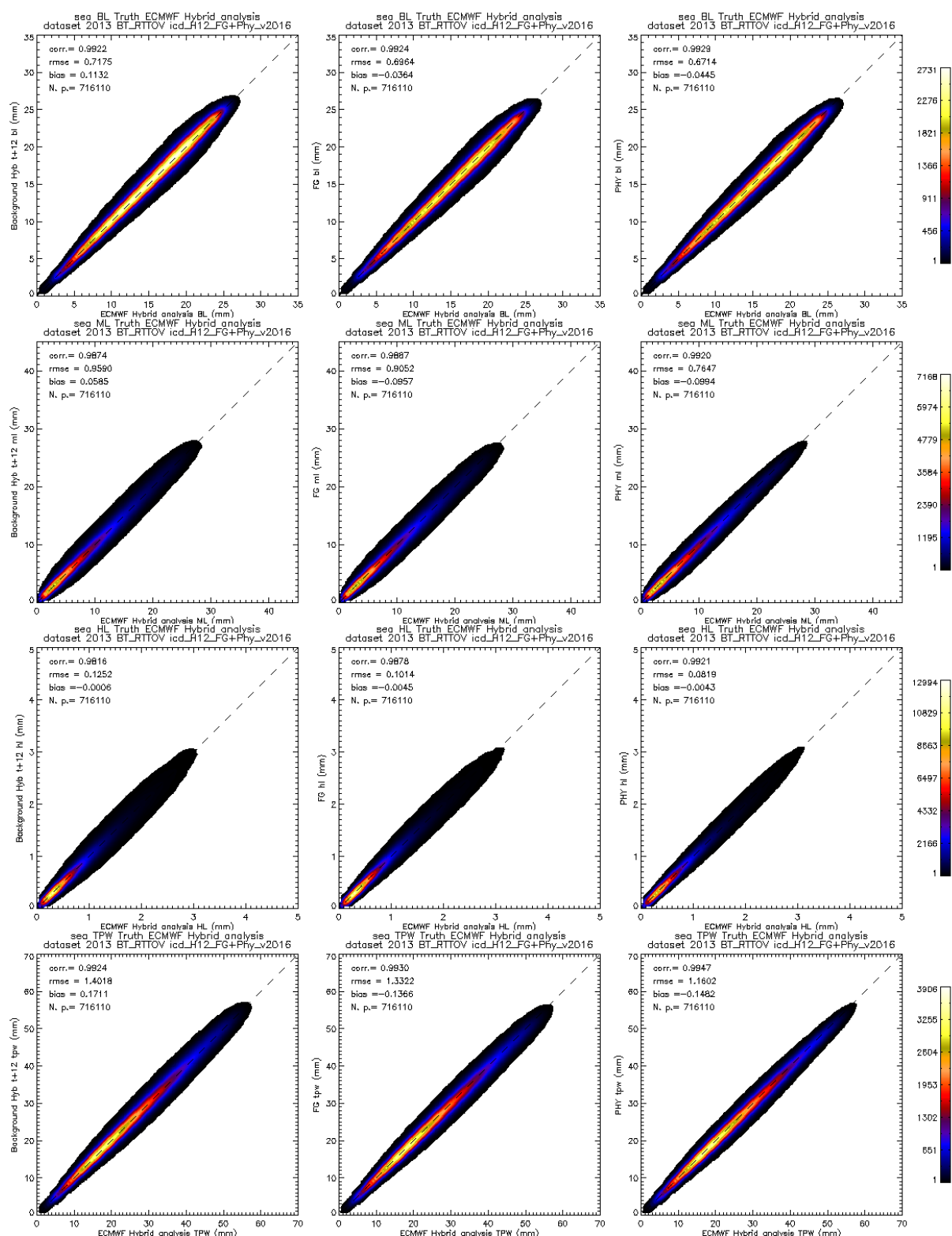
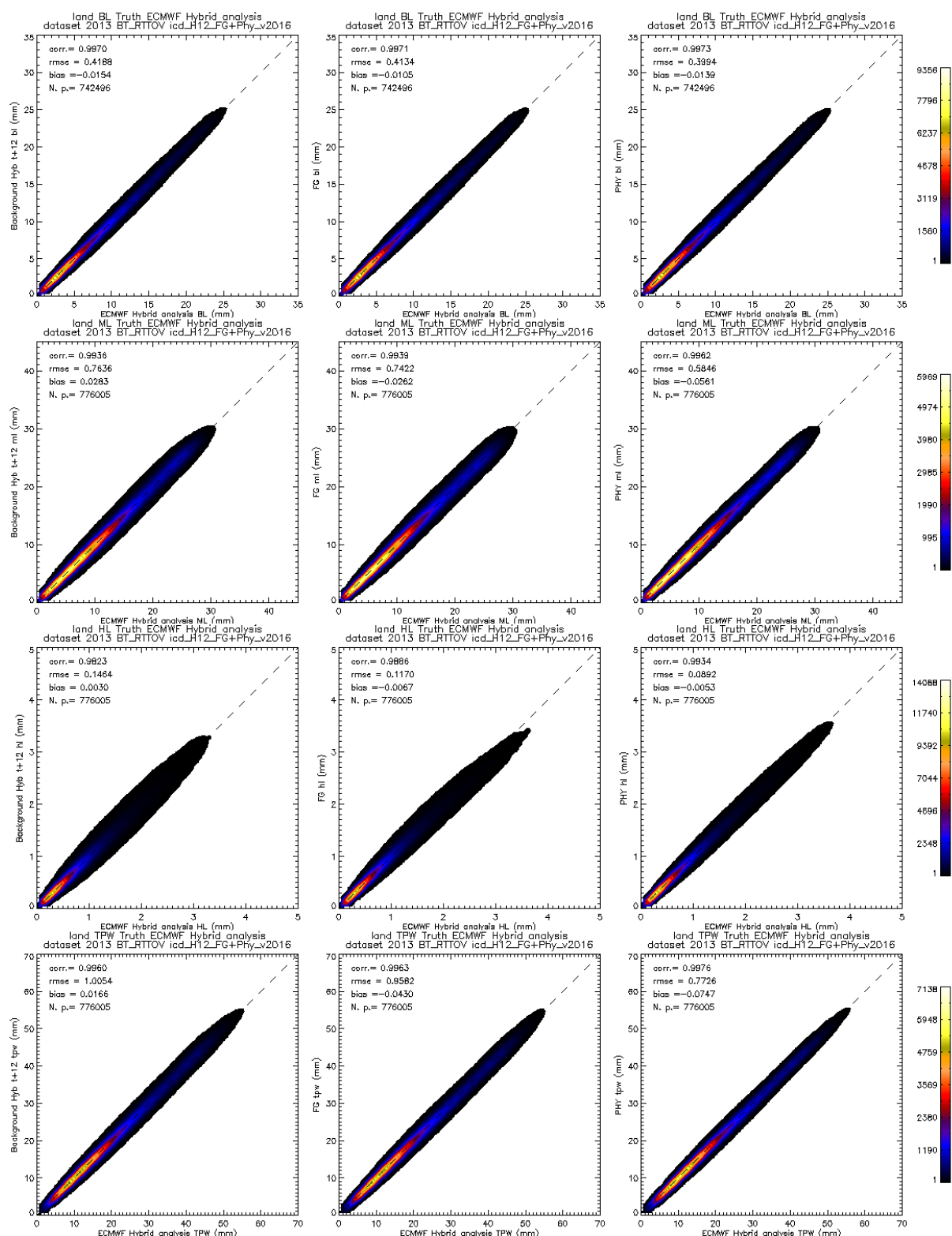


Figure 8: **BT\_RTTOV** case: LPW and TPW 2D histograms over sea validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input **BT\_RTTOV** (t+00), right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.



**Figure 9: BT\_RTTOV case: LPW and TPW 2D histograms over land validation points in odd pixels of year 2013.** From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input BT\_RTTOV (t+00), right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.



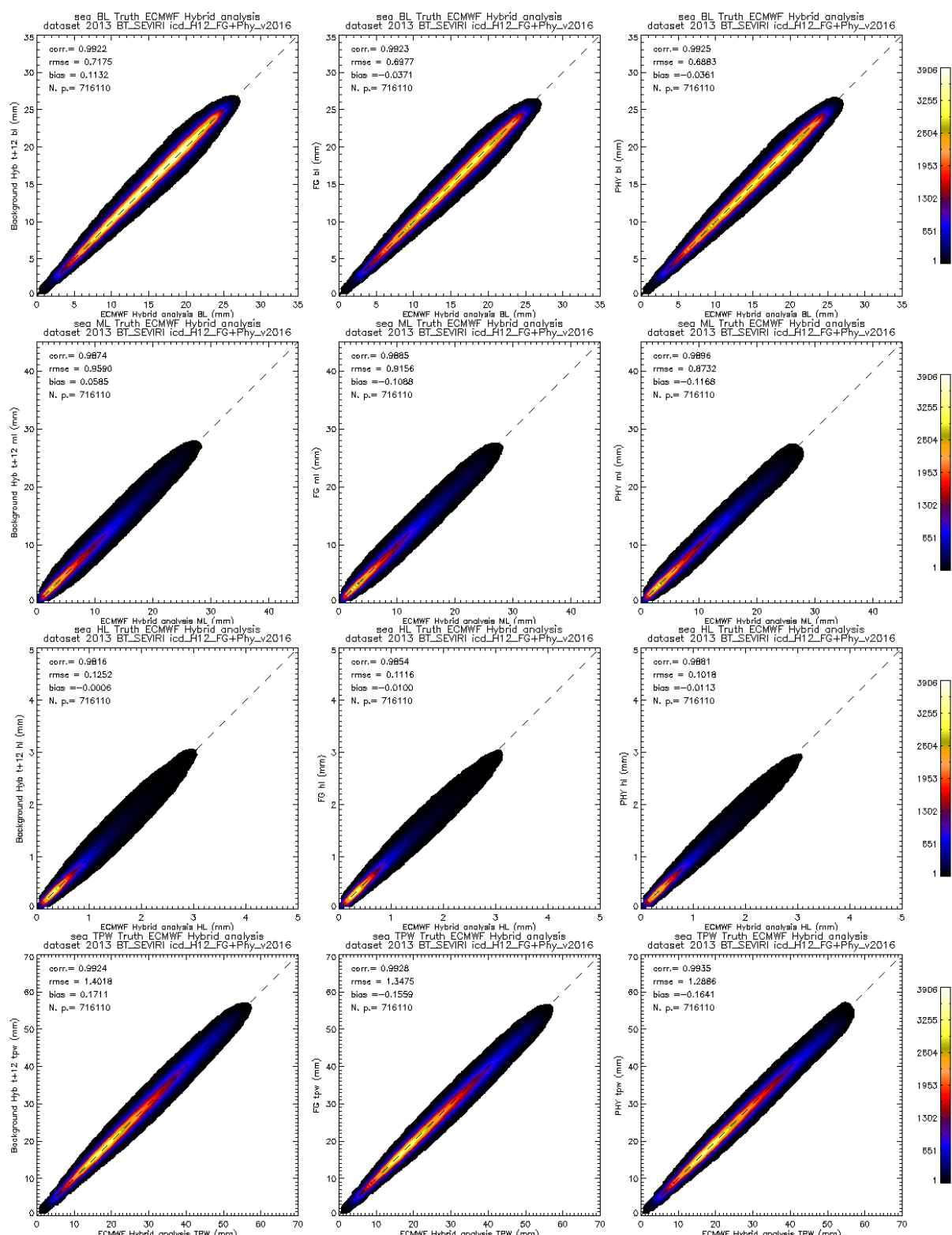


Figure 10: **BT\_SEVIRI** case: LPW and TPW 2D histograms over sea validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis(t+00) profiles.

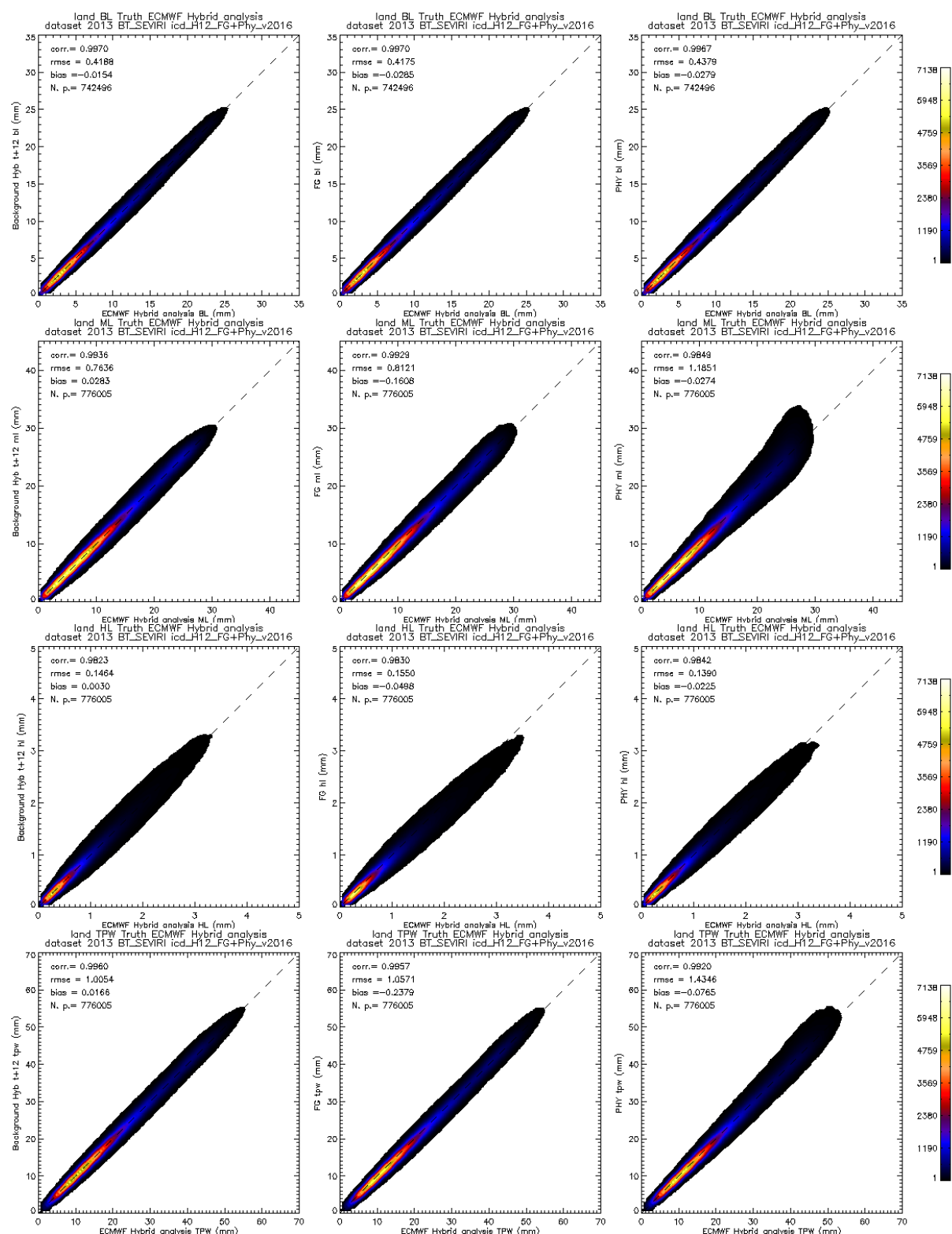
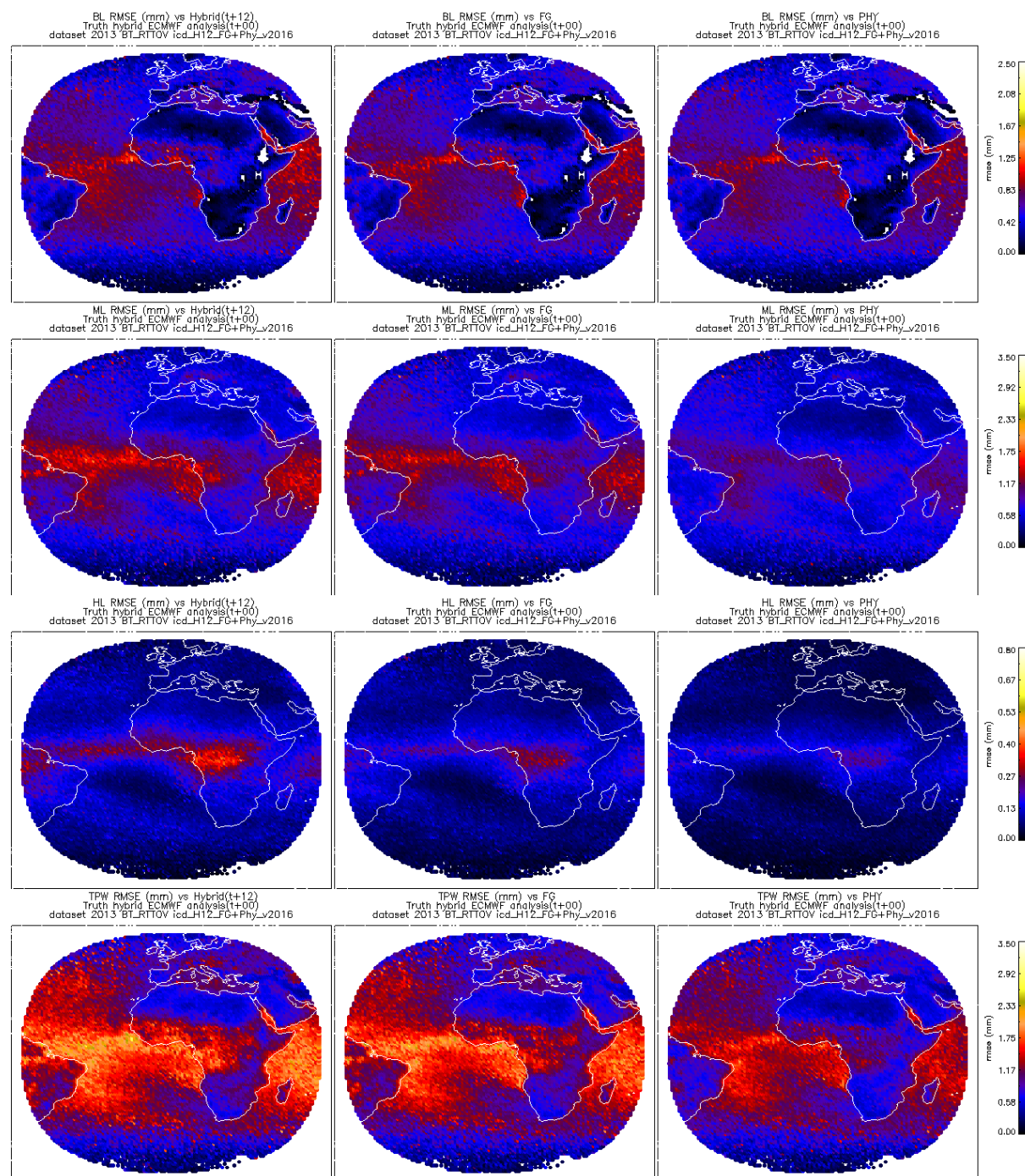


Figure 11: **BT\_SEVIRI** case: LPW and TPW 2D histograms over **land** validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.

### 3.4 SPATIAL ANALYSIS OF GEO ISHAI PARAMETERS

In Figures 12 and 14 the spatial performance of the LPW and TPW parameters for BT\_RTTOV case and BT\_SEVIRI case respectively are shown.



**Figure 12: BT\_RTTOV case: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.**



The greatest values of ML and HL RMSE appear near to the equatorial belt. But, when the relative ML RMSE are calculated, this effect disappears due to it has been caused by the high amount of precipitable water close to the equatorial belt. This effect can be seen in Figures 13 and 15.

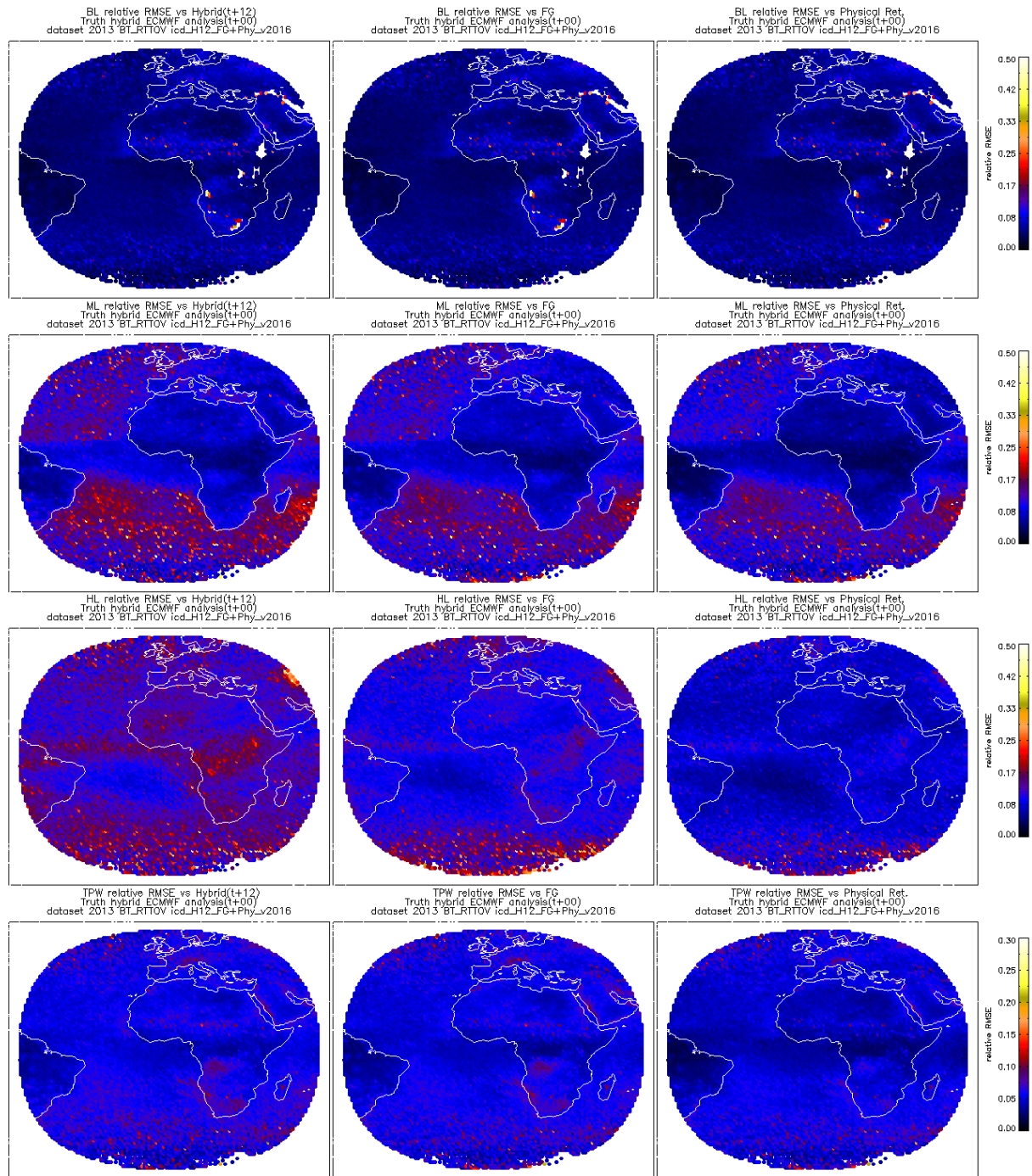
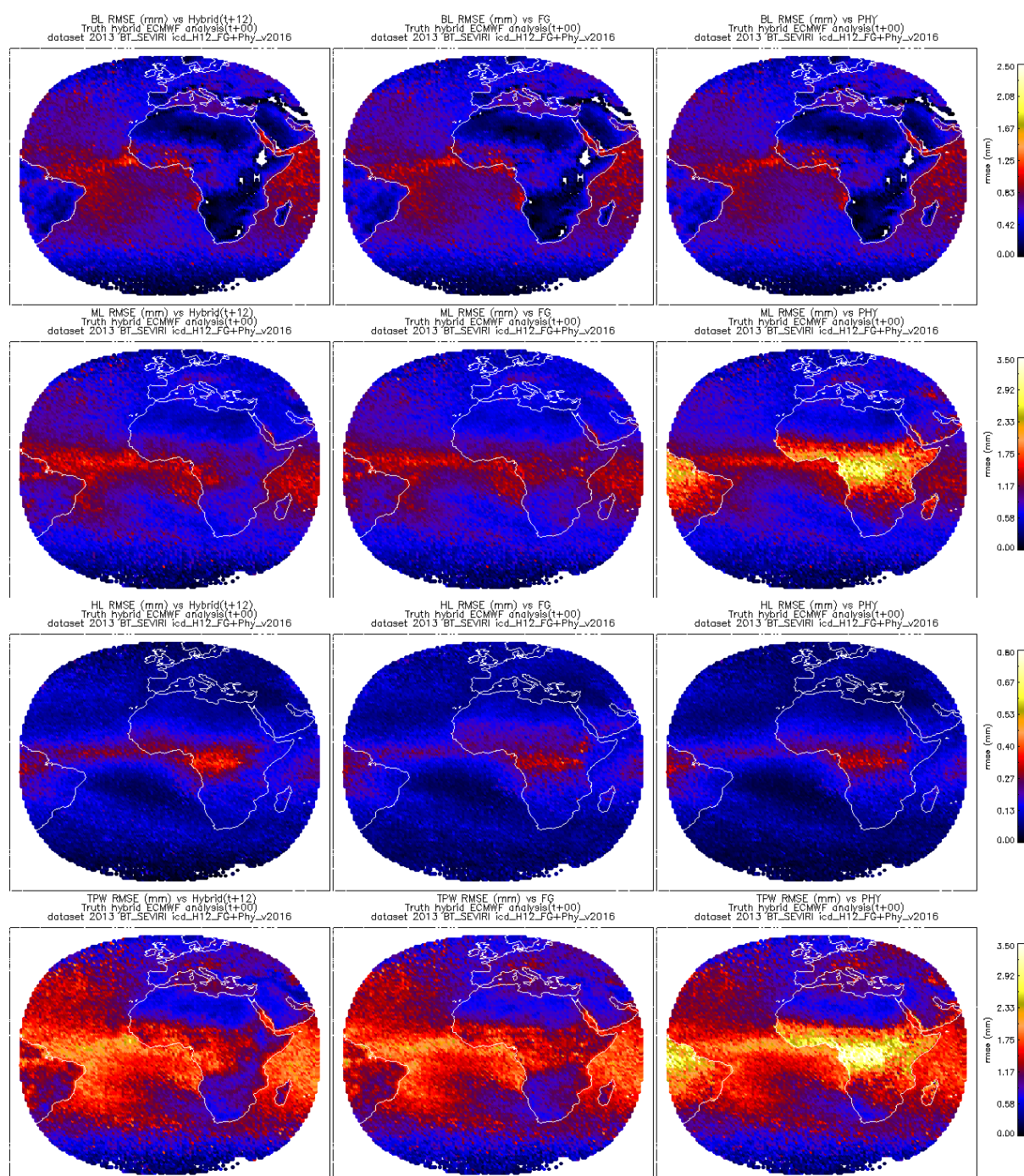


Figure 13: *BT\_RTTOV* case: Same that Figure 12 but relative RMSE instead of RMSE.



**Figure 14: BT\_SEVIRI case: Spatial distribution of the BL, ML, HL and TPW RMSE over validation points in 2013 for odd pixels dataset. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) BL, ML, HL and TPW RMSE calculated after FG step profile, right) BL, ML, HL and TPW RMSE calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWPHyb ECMWF analysis(t+00) profiles.**



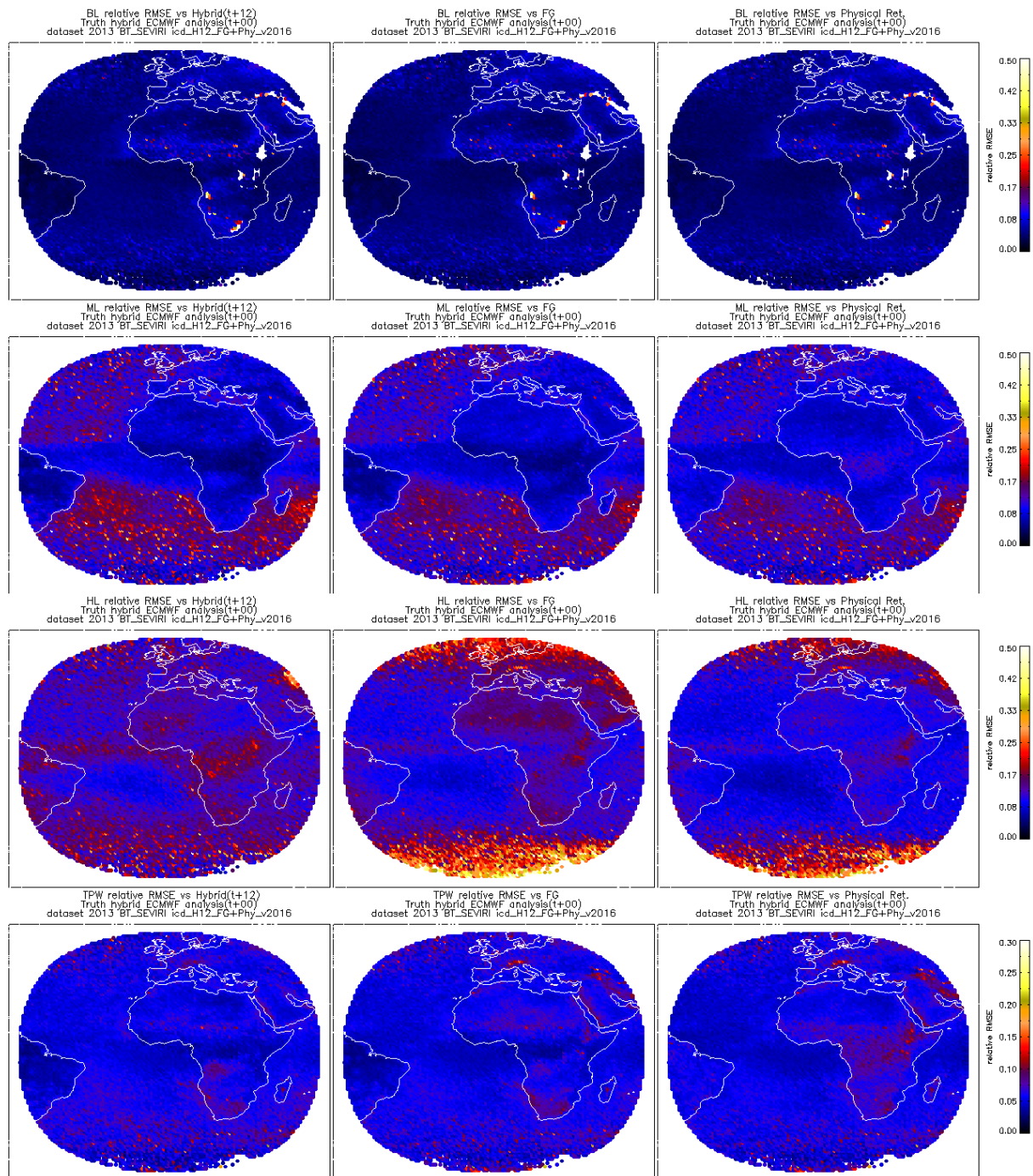


Figure 15: *BT\_SEVIRI* case: Same that Figure 14 but relative RMSE instead of RMSE.

### 3.5 STATISTICAL SUMMARY OF LPW AND STABILITY INDICES

In order to allow a better comparison, the statistical values that appear inside the 2D histograms have been collected below in Tables 3 to 6.

BL sea	NWPHyb(t+12)	FG	Phy. Retrieval	BL land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,717	0,696	0,671	RMSE (kg/m <sup>2</sup> )	0,419	0,413	0,399
BIAS (kg/m <sup>2</sup> )	0,113	-0,036	-0,045	BIAS (kg/m <sup>2</sup> )	-0,015	-0,010	-0,014
ML sea	NWPHyb(t+12)	FG	Phy. Retrieval	ML land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,959	0,905	0,765	RMSE (kg/m <sup>2</sup> )	0,764	0,742	0,585
BIAS (kg/m <sup>2</sup> )	0,059	-0,096	-0,099	BIAS (kg/m <sup>2</sup> )	0,028	-0,026	-0,056
HL sea	NWPHyb(t+12)	FG	Phy. Retrieval	HL land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,125	0,101	0,082	RMSE (kg/m <sup>2</sup> )	0,146	0,117	0,089
BIAS (kg/m <sup>2</sup> )	-0,001	-0,005	-0,004	BIAS (kg/m <sup>2</sup> )	0,003	-0,007	-0,005
TPW sea	NWPHyb(t+12)	FG	Phy. Retrieval	TPW land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	1,402	1,332	1,160	RMSE (kg/m <sup>2</sup> )	1,005	0,958	0,773
BIAS (kg/m <sup>2</sup> )	0,171	-0,137	-0,148	BIAS (kg/m <sup>2</sup> )	0,017	-0,043	-0,075

*Table 3: BT\_RTTOV case: Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.*

BL sea	NWPHyb(t+12)	FG	Phy. Retrieval	BL land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,717	0,698	0,688	RMSE (kg/m <sup>2</sup> )	0,419	0,417	0,438
BIAS (kg/m <sup>2</sup> )	0,113	-0,037	-0,036	BIAS (kg/m <sup>2</sup> )	-0,015	-0,029	-0,028
ML sea	NWPHyb(t+12)	FG	Phy. Retrieval	ML land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,959	0,916	0,873	RMSE (kg/m <sup>2</sup> )	0,764	0,812	1,185
BIAS (kg/m <sup>2</sup> )	0,059	-0,109	-0,117	BIAS (kg/m <sup>2</sup> )	0,028	-0,161	-0,027
HL sea	NWPHyb(t+12)	FG	Phy. Retrieval	HL land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,125	0,112	0,102	RMSE (kg/m <sup>2</sup> )	0,146	0,155	0,139
BIAS (kg/m <sup>2</sup> )	-0,001	-0,010	-0,011	BIAS (kg/m <sup>2</sup> )	0,003	-0,050	-0,022

BL sea	NWPHyb(t+12)	FG	Phy. Retrieval	BL land	NWPHyb(t+12)	FG	Phy. Retrieval
TPW sea	NWPHyb(t+12)	FG	Phy. Retrieval	TPW land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	1,402	1,348	1,289	RMSE (kg/m <sup>2</sup> )	1,005	1,057	1,435
BIAS (kg/m <sup>2</sup> )	0,171	-0,156	-0,164	BIAS (kg/m <sup>2</sup> )	0,017	-0,238	-0,077

**Table 4: BT\_SEVIRI case:** Statistical parameters for BL, ML, HL and TPW parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.

In the case of GEO-iSHAI validation with real SEVIRI BTs the performance has the same behaviour but with higher figures and irregular distribution of the RMSE over the land pixels. These irregularities in the figures of the statistical value over land are due to the issues explained in Section 3.1.

ML parameter shows a significant theoretical reduction in RMSE with GEO-iSHAI. From the values of Table 3 for sea pixels in the BT\_RTTOV case, a reduction in ML RMSE of 6% for FG step and 25% after physical retrieval can be expected over sea pixels. The reduction of ML RMSE over land pixels is greater than over sea pixels and it represents a 30% of reduction in the ML RMSE after the physical retrieval step.

After the inspection of Table 4 (BT\_SEVIRI case) the values using actual bias corrected SEVIRI BT are not so large; but there is still one reduction of the ML RMSE of 5% for FG step and 10% after physical retrieval in sea pixels. It should be remembered that these values have been obtained after the screening to remove the pixels with largest BT\_distance and BT\_RMS. If a perfect screening would be possible and a real truth could be obtained, then the performance would tend to the theoretical reduction in RMSE for the BT\_RTTOV case. As in the analysis of previous statistical values, the combinations of several sources of errors and uncertainties over land are a potential reason to explain the worse performance of ML over land.

In the case of HL parameter the percentage in the reduction with GEO-iSHAI module are even higher. For HL RMSE in the BT\_RTTOV case the theoretical reduction is around 50% over sea pixels and 60% over land pixels after physical retrieval step. In the BT\_SEVIRI case the reduction is round the half of the theoretical one (22% over sea pixels and 5% over land pixels). This better performance of HL parameter confirms that the WV channels have the greatest contribution and the source of errors as emissivity uncertainties and skin temperature affect less the HL parameter.

The former confirms the results of the vertical analysis of the performance made in Section 3.2 that showed a reduction in the RMSE and an improvement over the background NWP in the q profile at middle levels. The reduction of RMSE in the middle levels of the q profile is likely the main contribution for the reduction in the TPW RMSE.

In the case of the stability indices, there is not a clear statistical reduction in the RMSE with GEO iSHAI. Looking at Table 5 for the BT\_RTTOV case, the performance is better for the stability indices which involve as lower level the 850 hPa (Showalter Index and KI). Likely, this is due to the fact that SEVIRI has limited information to improve the temperature vertical information beyond the forecast. This fact is also explained from the vertical analysis of section 3.2; the highest reduction on q RMSE is on middle and high levels due to WV channels. Thus, it would be advisable to start looking for optimal stability indices from satellite retrievals of temperature and humidity profiles.

But although the statistical validation is not much better, the GEO-iSHAI stability indices have a great added value because SEVIRI provides useful spatial and temporal resolution. It is important

to take into account that a certain degree of disagreement between ECMWF analysis and actual temperature and humidity profiles always exists. For this reason, this is not always a negative aspect that the statistical values in BT\_SEVIRI case are greater than the ones in BT\_RTTOV case because it reflects the fact that actual SEVIRI BTs from the actual world are not the same that the synthetic and ideal RTTOV BTs ( $t+00$ ). One of the added values of the GEO-iSHAI is to show where and when there is a disagreement between ECMWF forecast or analysis against the actual bias corrected SEVIRI BTs retrieved profiles. Thus, the GEO-iSHAI stability parameters are able to delimitate the region where instability is growing before convection is triggered, as it can be seen on the study case loops or in the near real time loops in <http://www.nwcsaf.org>.

LI sea	NWPHyb(t+12)	FG	Phy. Retrieval	LI land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	0,640	0,632	0,629	RMSE (°C)	0,604	0,598	0,591
BIAS (°C)	-0,055	-0,008	-0,002	BIAS (°C)	0,071	0,026	0,030
SHW sea	NWPHyb(t+12)	FG	Phy. Retrieval	SHW land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	1,198	1,148	1,116	RMSE (°C)	0,655	0,659	0,644
BIAS (°C)	-0,156	0,047	0,052	BIAS (°C)	0,028	0,028	0,034
KI sea	NWPHyb(t+12)	FG	Phy. Retrieval	KI land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	3,439	3,311	3,160	RMSE (°C)	2,283	2,235	2,082
BIAS (°C)	0,250	-0,021	-0,011	BIAS (°C)	-0,025	-0,051	-0,064

*Table 5: **BT\_RTTOV** case: Statistical parameters for Lifted Index (LI), Showalter Index (SHW) and K Index (KI) parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.*

LI sea	NWPHyb(t+12)	FG	Phy. Retrieval	LI land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	0,640	0,633	0,635	RMSE (°C)	0,604	0,605	0,627
BIAS (°C)	-0,055	-0,003	-0,005	BIAS (°C)	0,071	0,030	0,028
SHW sea	NWPHyb(t+12)	FG	Phy. Retrieval	SHW land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	1,198	1,150	1,134	RMSE (°C)	0,655	0,673	0,726
BIAS (°C)	-0,156	0,048	0,049	BIAS (°C)	0,028	0,071	0,063
KI sea	NWPHyb(t+12)	FG	Phy. Retrieval	KI land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (°C)	3,439	3,323	3,227	RMSE (°C)	2,283	2,285	2,384
BIAS (°C)	0,250	-0,021	0,010	BIAS (°C)	-0,025	-0,231	-0,042

*Table 6: **BT\_SEVIRI** case: Statistical parameters for Lifted Index (LI), Showalter Index (SHW) and K Index (KI) parameters over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.*



### 3.6 VALIDATION OF GEO-iSHAI TOZ: TOTAL OZONE

In this Section the validation results of the Total Ozone (TOZ) are shown. GEO-iSHAI TOZ output has been introduced as a new output in release 2016. TOZ is calculated using only the non-linear regression steps; see the ATBD [RD.3].

In Figure 16 the ozone profile rmse between background NWP-Hyb  $t+12$  forecast and the analysis and the ozone profile rmse between the estimated GEO-iSHAI ozone and the analysis are shown.

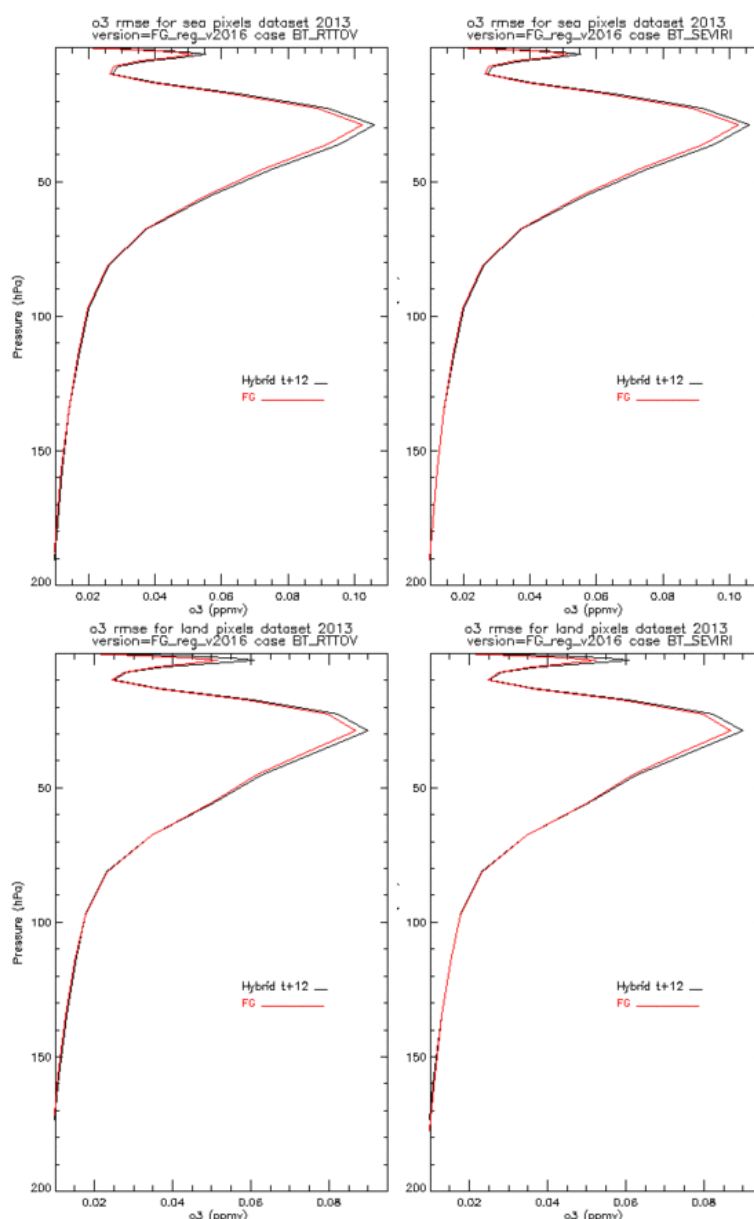


Figure 16: RMSE of ozone profiles (ppmv) at different steps compared with ECMWF analysis ( $t+00$ ) hybrid profiles. Left) BT\_RTTOV case, right) BT\_SEVIRI case. Top) RMSE of ozone over sea pixels, bottom) RMSE of ozone over land pixels. Odd pixels in 2013

In Figure 17 the spatial ozone rmse between background NWP-Hyb  $t+12$  forecast and the analysis and the ozone profile rmse between the estimated GEO-iSHAI ozone and the analysis are shown. There is not any region with degraded performance and BT\_SEVIRI and BT\_RTTOV spatial performance are similar.

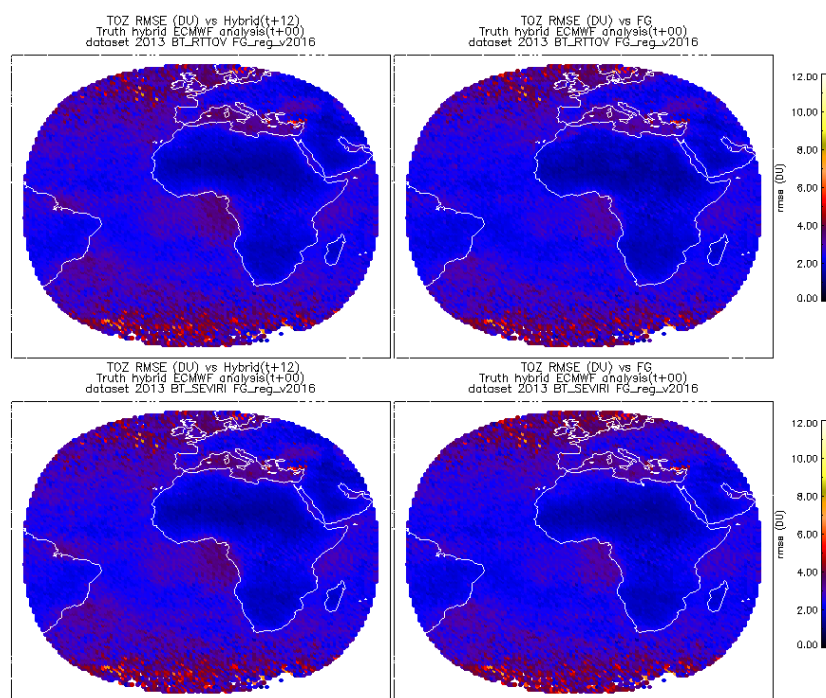


Figure 17: Spatial distribution of the TOZ RMSE over validation points in 2013 for odd pixels dataset. (top) BT\_RTTOV case (bottom) BT\_SEVIRI case. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB ( $t+12$ ), right) GEO-TOZ RMSE calculated after non-linear regression step profile. In all case the ground truth are TOZ calculated from NWP-Hyb ECMWF analysis( $t+00$ ) profiles.

In Figures 18 and 19 the two dimensional histograms for TOZ between background NWP-Hyb  $t+12$  forecast and the analysis and between the estimated GEO-iSHAI TOZ and the analysis TOZ are shown. It can be seen that the two dimensional histograms of TOZ from BT\_SEVIRI case are similar to the ones for BT\_RTTOV case and this confirms the results of Figures 15 and 16.

The statistical values (RMSE, bias and correlation) that appear in the 2D histograms are also written at the end of this Section 3.5 for a more comfortable read and comparison in Tables 7 and 8.

The reductions on TOZ rmse in the BT\_RTTOV case are of 5% and 2% over sea and land pixels respectively. In the SEVIRI case a reduction of a 4% is retained over sea pixels and an increase of 1% over land. Other time, this represents the difference between synthetic RTTOV case and real SEVIRI case.

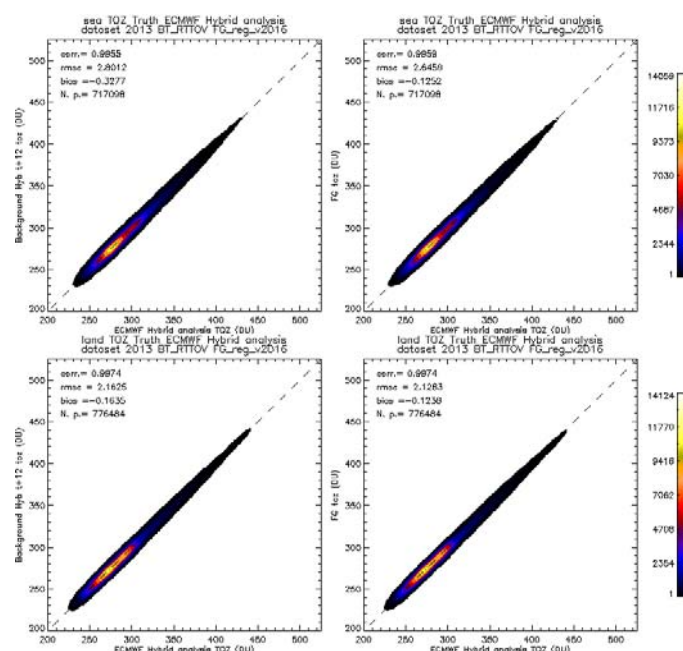


Figure 18: **BT\_RTTOV** case TOZ 2D histograms over odd pixels of year 2013. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background t+12ECMWF hybrid GRIB), (right) GEO-TOZ RMSE calculated after non-linear regression step profile.

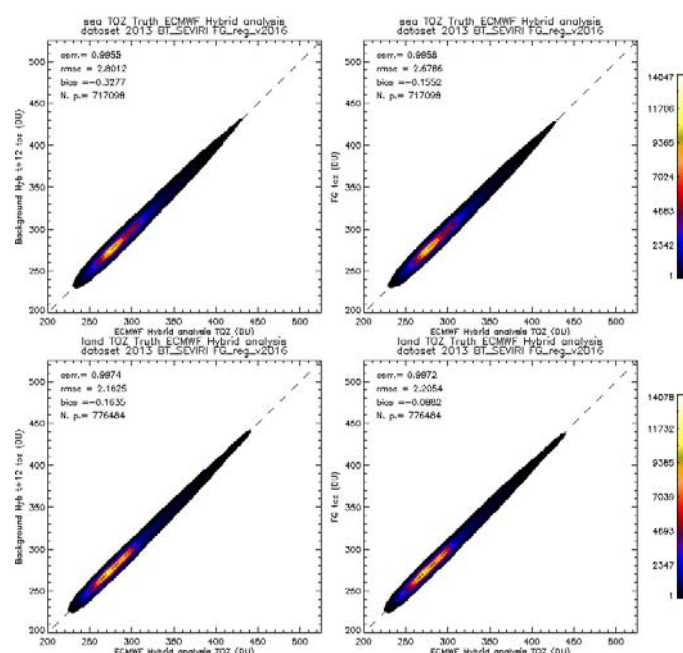


Figure 19: **BT\_RTTOV** case TOZ 2D histograms over odd pixels of year 2013. (top) sea TOZ. (bottom) land TOZ. Left) TOZ RMSE calculated directly from background t+12ECMWF hybrid GRIB), (right) GEO-TOZ RMSE calculated after non-linear regression step profile.

TOZ sea	NWPHyb(t+12)	FG
RMSE (DU)	2,801	2,646
BIAS (DU)	-0,328	-0,125

TOZ land	NWPHyb(t+12)	FG
RMSE (DU)	2,163	2,128
BIAS (DU)	-0,164	-0,124

Table 7: **BT\_RTTOV** case: Statistical parameters for Total Ozone (TOZ) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.

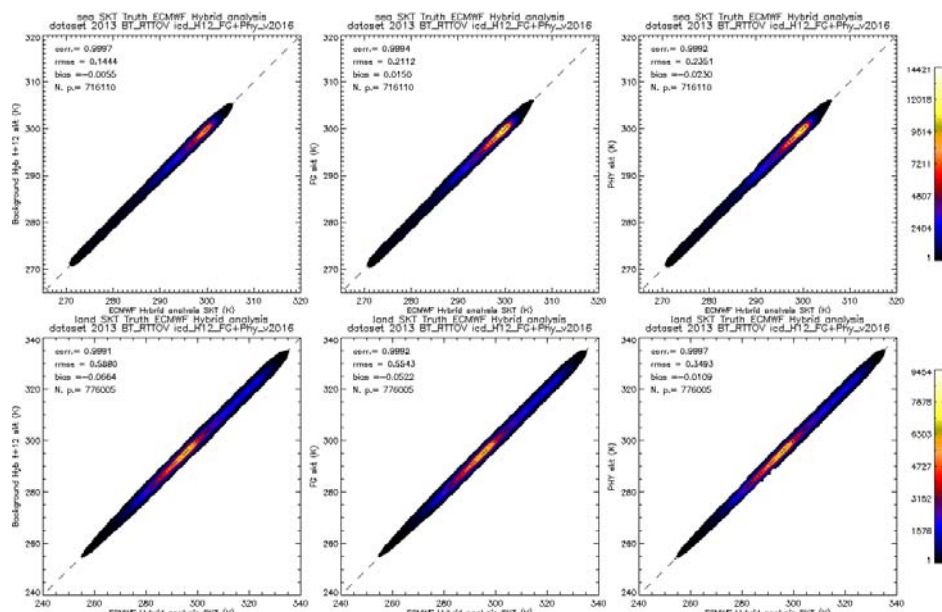
TOZ sea	NWPHyb(t+12)	FG	TOZ land	NWPHyb(t+12)	FG
RMSE (DU)	2,801	2,679	RMSE (DU)	2,163	2,205
BIAS (DU)	-0,328	-0,156	BIAS (DU)	-0,164	-0,088

**Table 8: BT\_SEVIRI case:** Statistical parameters for Total Ozone (TOZ) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.

### 3.7 VALIDATION OF GEO-ISHAI SKT: SKIN TEMPERATURE

In this Section the validation results of the Skin Temperature (SKT) are shown. This GEO-iSHAI output has been introduced as a new output in release 2016. Below, the Tables and Figures with the statistical parameters for the SKT parameters are shown. SKT is written just for nowcasting purposes and in order for the users to have access to this parameter. As an example, as SKT is used in the RTTOV calculations, the inspection of spatial gradients and temporal tendency could be used to detect the presence of non-adequately detected clouds or errors in the background NWP SKT.

The SKT should be taken as an indicative output and it should not be considered as SST or LST products because more controls and spatial and temporal tests would be needed. The SKT field of ECMWF has not a great quality especially over land pixels; due to this fact the spatial rmse in BT\_SEVIRI case show great values over land in Figure 22. The result of the Figures and Tables in this SKT Section is that GEO-iSHAI SKT could be used to inform users of the discrepancies between the background NWP SKT and one optimal SKT in the pixels; but it must be taken into account that the discrepancies could be due to physical reasons, due to undetected clouds or due to error in the emissivities, etc.



**Figure 20: BT\_RTTOV case SKT 2D histograms over odd pixels of year 2013.** (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) SKT RMSE calculated after FG step profile, right) SKT RMSE calculated after



physical retrieval step profile. In all case the ground truth is SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles.

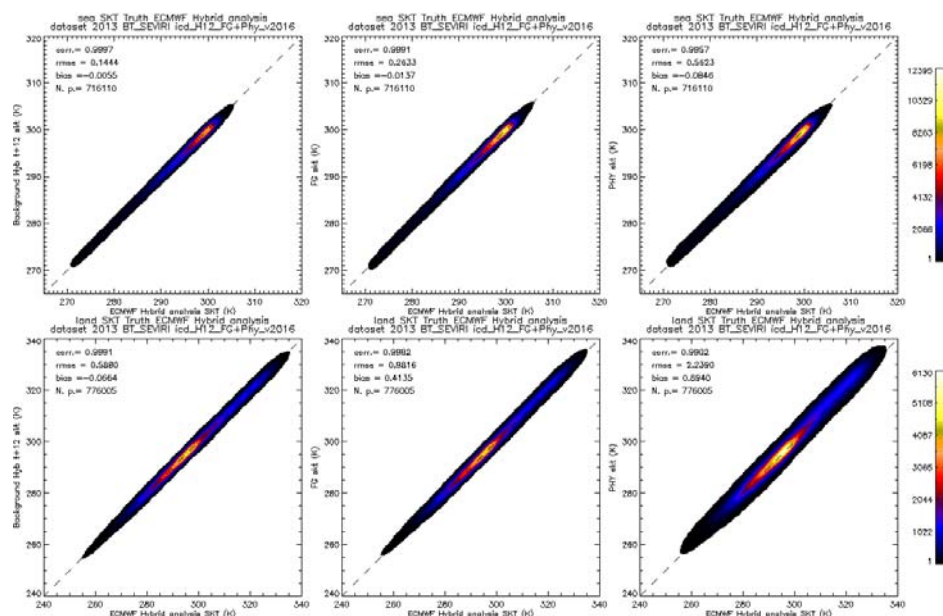


Figure 21: BT\_SEVIRI case SKT 2D histograms over odd pixels of year 2013. (top) sea SKT. (bottom) land SKT. Left) SKT RMSE calculated directly from background ECMWF hybrid GRIB (t+12), centre) SKT RMSE calculated after FG step profile, right) SKT RMSE calculated after physical retrieval step profile. In all case the ground truth is SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles.

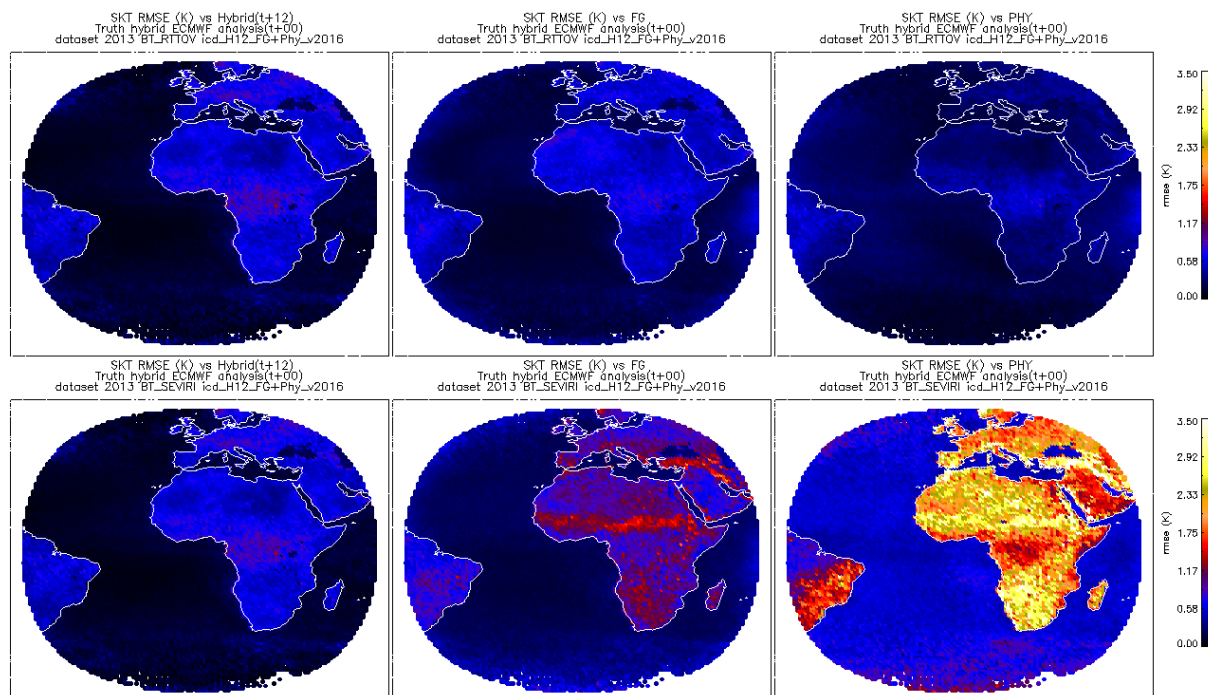


Figure 22: Spatial distribution of the SKT RMSE over validation points in 2013 for odd pixels dataset. (top) BT\_RTTOV case (bottom) BT\_SEVIRI case. Left) TOZ RMSE calculated directly from background ECMWF hybrid GRIB (t+12), (centre) SKT RMSE calculated after FG non-linear



regression step,( right) GEO-SKT RMSE calculated after physical retrieval step. In all case the ground truth are SKT calculated from NWP-Hyb ECMWF analysis(t+00) profiles.

SKT sea	NWPHyb(t+12)	FG	Phy. Retrieval	SKT land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (K)	0,144	0,211	0,235	RMSE (K)	0,588	0,554	0,349
BIAS (K)	-0,005	0,015	-0,023	BIAS (K)	-0,066	-0,052	-0,011

**Table 9: BT\_RTTOV case:** Statistical parameters for Skin Temperature (SKT) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.

SKT sea	NWPHyb(t+12)	FG	Phy. Retrieval	SKT land	NWPHyb(t+12)	FG	Phy. Retrieval
RMSE (K)	0,144	0,263	0,562	RMSE (K)	0,588	0,982	2,239
BIAS (K)	-0,005	-0,014	-0,085	BIAS (K)	-0,066	0,413	0,894

**Table 10: BT\_SEVIRI case:** Statistical parameters for Skin Temperature (SKT) parameter over the Full Disk validation points in year 2013 for odd pixels dataset. Left) sea pixels, right) land pixels.

## 4. CONCLUSIONS

After this validation of GEO-iSHAI, some conclusions can be obtained:

- To build a good training and validation dataset is a very important and standing task for us. This continuous task has allowed the generation of a huge dataset. The use of this dataset has two main aims. The first one is to serve for the validation and tuning of the current version of the algorithm. The second one is the training, testing and validation of new versions of GEO-iSHAI. In the 2016 version all the coefficients have been calculated from the GEO-iSHAI validation and training datasets (using ECMWF profiles).
- It is needed to improve the process to screen out cloud contaminated pixels. As can be seen in Figure 25 there are still in version 2016 cloudy contaminated pixels that increase the rmse values on the case of real SEVIRI tests.
- The GEO-iSHAI processor with coefficients trained using the ECMWF NWP Hybrid dataset, allowing the use of ECMWF GRIB files on hybrid levels, improves strongly the performance over the regular version of the GEO-iSHAI. For the users that use ECMWF to feed NWC SAF software in real time, it is strongly recommend to make the effort of downloading from the MARS the ECMWF GRIB files on hybrid levels and to use the GEO-iSHAI HYB mode. For the users that use their own NWP models to feed NWC SAF software in real time, it is strongly recommended they make the effort to provide optimal vertical resolution to GEO-iSHAI.
- The performance of the regular GEO iSHAI version 2016 with fixed pressure NWP input will tend to be similar than the one for the GEO-iSHAI hybrid mode when the number of available vertical levels increases. Then, it is strongly recommend to the users to feed to the regular GEO iSHAI inputs with the highest available number of fixed pressure levels.
- It is also recommended to feed the GEO-iSHAI with the highest possible temporal and spatial resolution.
- Validation has been performed for the complete SEVIRI disc.
- Best results are obtained for humidity in medium layers due to the contribution of the two water vapor channels. In this layer the GEO-iSHAI improves the information beyond the background NWP on the humidity profile.
- SEVIRI has limited information to improve the vertical information beyond the forecast, but it does provide useful spatial and temporal information. This limitation is clear for the vertical information of temperature and stability indices.
- The RMSE of the GEO iSHAI parameters are excellent (see Table 11 below) and all the parameters are better than is requested in the Product Requirement Document [AD.4]
- A good BT bias correction is essential for GEO-iSHAI. A mechanism to calculate and distribute frequently and updated SEVIRI BT bias correction has been implemented through a web page in the NWC SAF web, to provide frequent and rapid updates of the SEVIRI BT bias correction.
- Validation has been performed for an extended period of a complete year 2013. But there are more years available. During CDOP-2, a web page should be created in order to maintain updated and detailed validation documentation, examples, repository of case studies, etc.
- Table 11 summarizes the objective validation results in terms of RMSE; it summarizes the statistical values reported along the different sections of the document but the most important is that these figures represent a reduction in the RMSE from the background NWP, that on synthetic case could be greater than 50% for HL layer and 25 % for ML layer.
- This reduction of RMSE is lower using actual SEVIRI BTs but the algorithm is sensible to discrepancy between synthetic and actual SEVIRI BTs and then is able to advice the users of

any discrepancy in the forecast NWP used as background NWP; it is recommend the use of the parameters with the differences between the NWP used as input to GEO iSHAI and the retrieved profiles in order to detect any discrepancy between the forecast model and the actual observations of the SEVIRI images.

GEO iSHAI V3.0 summary of validation Results	Precipitable Water Low Layer – BL RMSE	Precipitable Water Medium Layer ML RMSE	Precipitable Water High Layer HL RMSE	Precipitable Water Total TPW RMSE	Showalter Index SHW RMSE
Against ECMWF Analysis – Over Sea Full Disk validation	0,688 (kg/m <sup>2</sup> )	0,873 (kg/m <sup>2</sup> )	0,102 (kg/m <sup>2</sup> )	1,289 (kg/m <sup>2</sup> )	1,134 (K)
Against ECMWF Analysis – Over Land Full Disk validation	0,438 (kg/m <sup>2</sup> )	1,185 (kg/m <sup>2</sup> )	0,139 (kg/m <sup>2</sup> )	1,435 (kg/m <sup>2</sup> )	0,726 (K)

*Table 11: Summary of the GEO iSHAI statistical parameters in 2013 for odd pixels of GEO iSHAI validation dataset using as input to GEO iSHAI actual SEVIRI BTs bias corrected.*

The statistical parameter of the above tables could be compared with the statistical accuracy values defined in the Product Requirement Table (PRT). The PRT values for GEO-iSHAI has been copied in the Table 12.

PRT rms for sea pixels	TPW (mm)	BL (mm)	ML (mm)	HL (mm)	Lifted Index (°C)	Showalter Index (°C)	K Index	SKT (K)	TOZ (DU)
Threshold Accuracy	3.5	2.5	2.5	0.5	3.0	3.0	6	4	20
Target Accuracy	1.9	1.0	1.7	0.2	1.5	1.5	3.5	2.5	9
Optimal Accuracy	1.2	0.7	0.8	0.1	1.0	1.1	3.25	1.5	7.5

*Table 12: Statistical accuracy values defined in the Product Requirement Table.*

## 5. ANNEX I: VALIDATION STATISTICS OVER EUROPE REGION.

Here are collected the statistical summary of LPW and stability indices over land on Europe region to show that the differences in the performance on RTTOV case and SEVIRI case are due to the issues exposed in the Section 3.1.

It has been considered as Europe region the land pixels with latitude greater than 36° N on the iSHAI validation set of 13001 validation pixels.

First the 2D histograms over land for LPW and TPW on RTTOV and SEVIRI cases on Europe are shown in Figures 23 and 24. In order to allow a better comparison, the statistical values that appear inside the 2D histograms have been collected below in Table 13.

As can be seen from the comparison of 2D histograms for land Europe cases (Figures 23 and 24) with the ones for the full disc (Figures 9 and 11), the tilt and spread on high LPW values (mainly on ML parameter) is not present and the 2D histograms are linear always.

As can be seen from the comparison of Table 13 with the Tables 3 and 4, in the case of GEO-iSHAI validation on Europe region over land pixels the performance is better than in the full disc validation. It is also present one increase in the rmse between RTTOV and SEVIRI cases but with lower rmse increment between RTTOV and SEVIRI case. This is explained due that over Europe region the desert and tropics issues are not present.

In Table 13 it has been added two columns (red columns) with the rmse and bias in tests made using real uncorrected bias SEVIRI BT as input to GEO-iSHAI. It can be seen that if real SEVIRI BTs are not bias BT corrected all the statistical figures are worse.

From Table 13 it is also possible to conclude that:

- the algorithm is able to make correct changes in the profile as can be seen in the RTTOV case statistical figures
- the algorithm is stable since the use of real uncorrected bias SEVIRI BT introduces greater error but the rmse does not increase uncontrolled
- since the only difference between the three case are the BTs input the increase of rmse are due to errors in not well filtered real SEVIRI BTs, emissivities issues and lack of good truth.
- the remaining levels of rmse greater on SEVIRI cases than RTTOV one are likely caused by the remaining undetected cloudy pixels inside the validations dataset. As example it can be seen in Figure 25 red pixels in the neighbourhood of cloudy pixels with large and opposite error in ML parameter in the neighbourhood of cloudy pixels. The red pixels are caused by undetected clouds or cloud contamination; in real time operations these pixels are screened out by forecasters but they are one source of great errors when calculating statistical parameters.

In future the validation will be repeated with other truth sources as radiosoundings or ground GPS receivers.

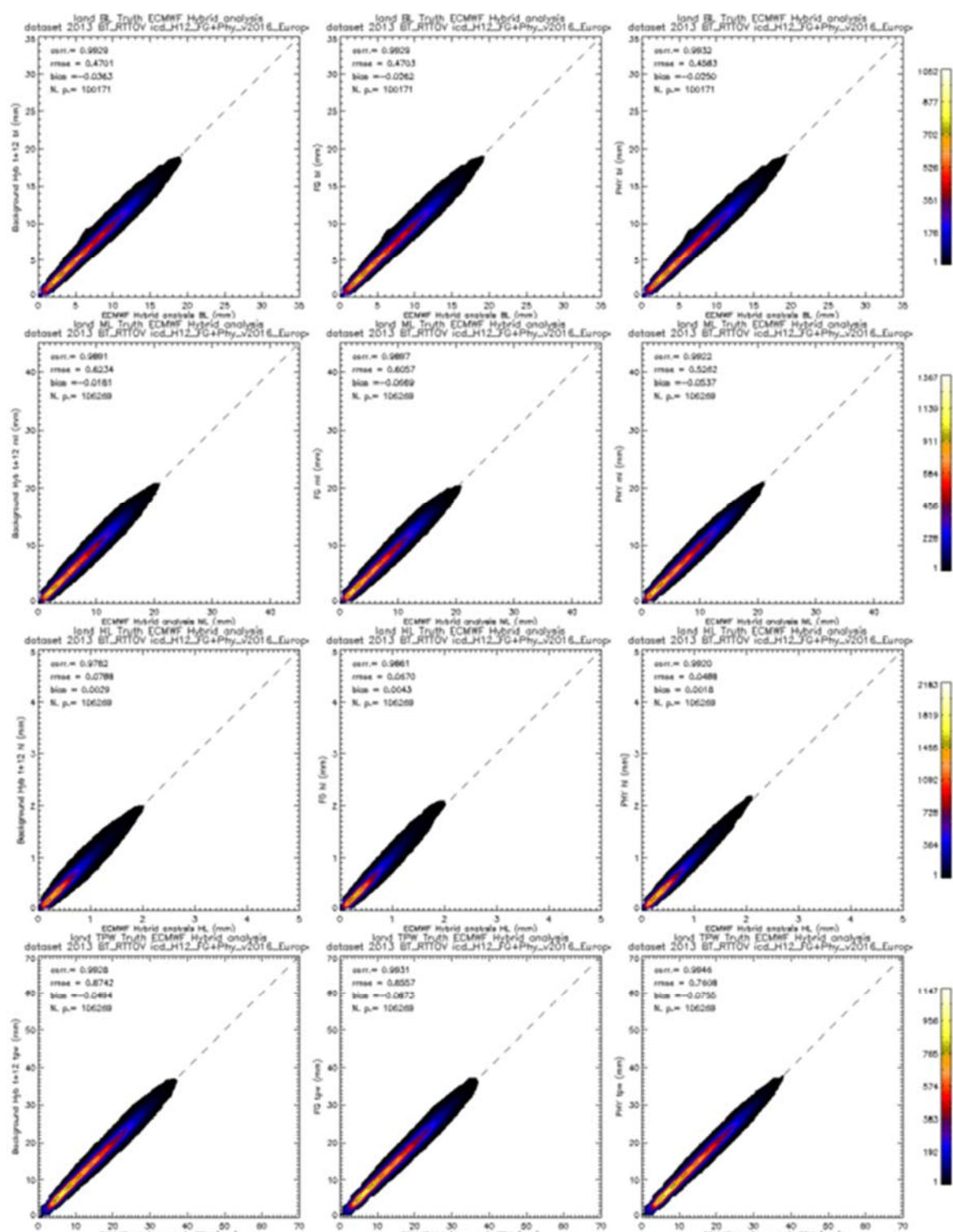


Figure 23: **BT\_RTTOV** case on Europe region: LPW and TPW 2D histograms over land on Europe region validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.



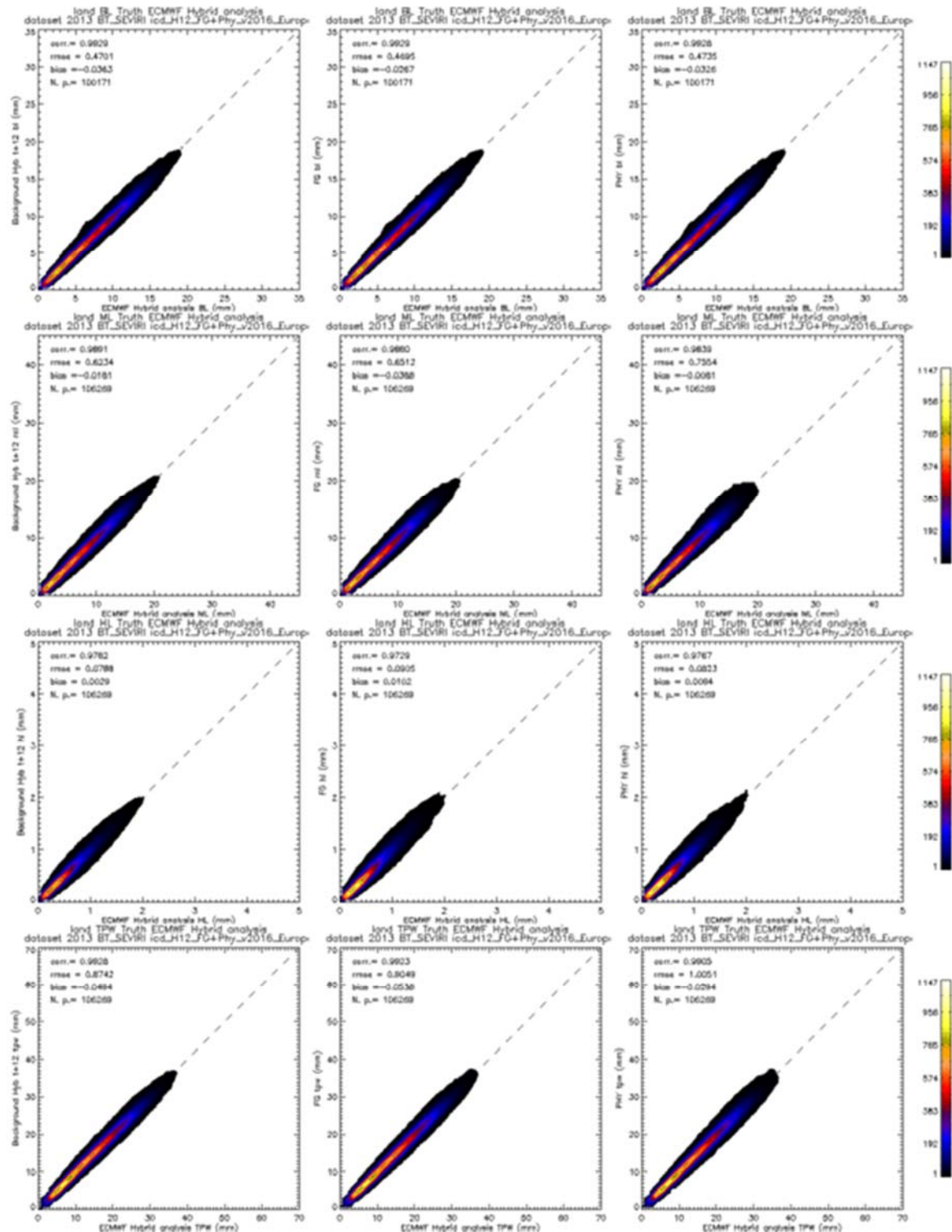
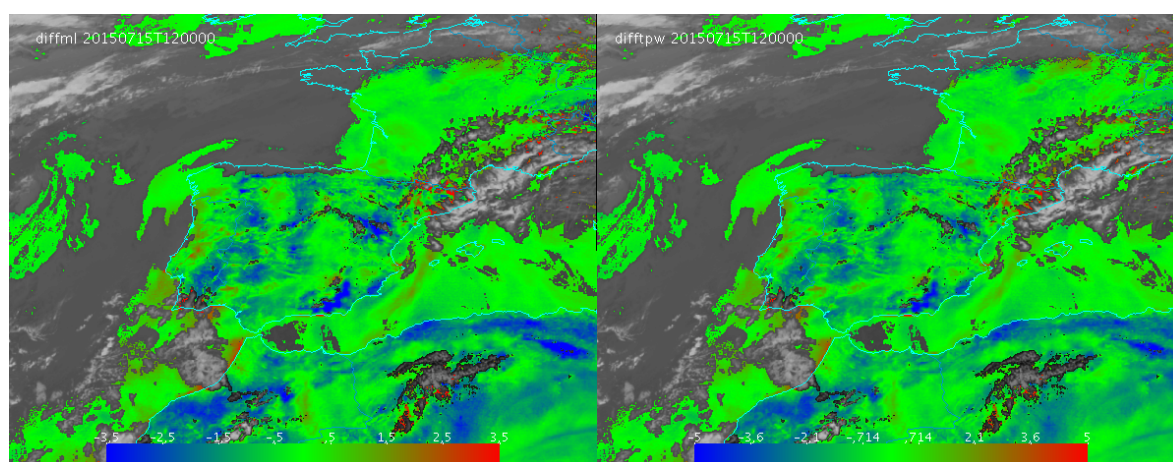


Figure 24: **BT\_SEVIRI** case on Europe region: LPW and TPW 2D histograms over **land** on Europe region validation points in odd pixels of year 2013. From top to bottom BL, ML, HL and TPW parameters. Left) BL, ML, HL and TPW parameters calculated directly from background ECMWF from hybrid profiles from (t+12) forecast, centre) BL, ML, HL and TPW parameters calculated after FG step profile using as input using actual bias corrected SEVIRI BT, right) BL, ML, HL and TPW parameters calculated after physical retrieval step profile. In all case the ground truth are the BL, ML, HL and TPW calculated from NWP-Hyb ECMWF analysis (t+00) profiles.

BL land	NWPHyb(t+12)	BT_RTTOV case FG	BT_RTTOV case Phy. Retrieval	BT_SEVIRI case FG	BT_SEVIRI case Phy. Retrieval	BT_SEVIRI_unc case FG	BT_SEVIRI_unc case Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,470	0,470	0,458	0,469	0,473	0,470	0,486
BIAS (kg/m <sup>2</sup> )	-0,036	-0,026	-0,025	-0,027	-0,033	-0,044	-0,023
ML land	NWPHyb(t+12)	BT_RTTOV case FG	BT_RTTOV case Phy. Retrieval	BT_SEVIRI case FG	BT_SEVIRI case Phy. Retrieval	BT_SEVIRI_unc case FG	BT_SEVIRI_unc case Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,623	0,606	0,526	0,651	0,755	0,698	0,793
BIAS (kg/m <sup>2</sup> )	-0,018	-0,067	-0,054	-0,039	-0,008	-0,232	0,057
HL land	NWPHyb(t+12)	BT_RTTOV case FG	BT_RTTOV case Phy. Retrieval	BT_SEVIRI case FG	BT_SEVIRI case Phy. Retrieval	BT_SEVIRI_unc case FG	BT_SEVIRI_unc case Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,079	0,067	0,049	0,090	0,082	0,119	0,085
BIAS (kg/m <sup>2</sup> )	0,003	0,004	0,002	0,010	0,009	-0,070	-0,029
TPW land	NWPHyb(t+12)	BT_RTTOV case FG	BT_RTTOV case Phy. Retrieval	BT_SEVIRI case FG	BT_SEVIRI case Phy. Retrieval	BT_SEVIRI_unc case FG	BT_SEVIRI_unc case Phy. Retrieval
RMSE (kg/m <sup>2</sup> )	0,874	0,856	0,761	0,905	1,005	0,978	1,042
BIAS (kg/m <sup>2</sup> )	-0,049	-0,087	-0,075	-0,054	-0,029	-0,344	0,006

*Table 13: Statistical parameters for BL, ML, HL and TPW parameters over land Europe validation points in year 2013 for odd pixels dataset. Blue column) Green columns) BT\_RTTOV case, light yellow columns) BT\_SEVIRI case, red columns) uncorrected bias BT BT\_SEVIRI case.*



*Figure 25: Example of GEO-iSHAI diff\_ML and diff\_TPW from 12 UTC on 15 July 2015 produced from SEVIRI on MSG-3. Red pixels (large error in ML parameter) in the neighbourhood of cloudy pixels are caused by undetected clouds or cloud contamination.*